

03-27-00

7

03/23/00

Customer No. 20350
 TOWNSEND and TOWNSEND and CREW LLP
 200 Embarcadero Center, 8th Floor
 San Francisco, California 94111-3834
 (415) 576-0200

Attorney Docket No. 19496-003020US

"Express Mail" Label No. EL170263749US

Date of Deposit: March 23, 2000

I hereby certify that this is being deposited with the United States Postal Service "Express Mail Post Office to Addressee" service under 37 CFR 1.10 on the date indicated above, addressed to:

Assistant Commissioner for Patents
 Washington, D.C. 20231

By: 



ASSISTANT COMMISSIONER FOR PATENTS
 U.S. PATENT APPLICATION
 Washington, D.C. 20231

Sir:

Transmitted herewith for filing under 37 CFR 1.53(b) is the

- ☒ patent application of
☐ continuation patent application of
☐ divisional patent application of
☐ continuation-in-part patent application of

Inventor(s)/Applicant Identifier: Stephen Eisenberg, Casey Case, George Cox III, Andrew Jamieson, Edward Rebar and Qiang Lui.

For: ZINC FINGER PROTEIN COMPOSITIONS

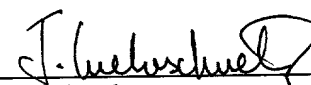
- ☒ This application claims priority from each of the following Application Nos./filing dates:
 60/126,238 filed March 24, 1999, 60/126,239 filed March 24, 1999, 60/146,596 filed July 30, 1999 and 60/146,615 filed July 30, 1999, the disclosures of which are incorporated by reference.
- ☐ Please amend this application by adding the following before the first sentence: "This application is a ☐ continuation ☐ continuation-in-part of and claims the benefit of U.S. Provisional Application No. 60/_____, filed _____, the disclosure of which is incorporated by reference."

Enclosed are:

- ☒ 55 Pages of Specification, 2 Pages of Claims, 1 Page of Abstract and 1 Sheet of Drawing;
☒ Unsigned Declaration and Power of Attorney;
☒ Return Postcard.

**In view of the Unsigned Declaration as filed with this application and pursuant to 37 CFR §1.53(f),
 Applicant requests deferral of the filing fee until submission of the Missing Parts of Application.**

DO NOT CHARGE THE FILING FEE AT THIS TIME.


 Joe Liebeschuetz

Reg No.: 37,505
 Attorneys for Applicant

Telephone:
 (415) 576-0200

Facsimile:
 (415) 576-0300

PA 3059297 v1

Inventor(s):

Qiang Lui
A Citizen of the United States of America
55 Williams Lane
Foster City
CA 94404

Assignee:

Entity: SMALL

TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834
(650) 326-2400

ZINC FINGER PROTEIN COMPOSITIONS

5 CROSS-REFERENCES TO RELATED APPLICATIONS

The present application claims priority to U.S. provisional applications 60/126,238, filed March 24, 1999, 60/126,239 filed March 24, 1999, 60/146,596 filed July 30, 1999 and 60/146,615 filed July 30, 1999, all of which are incorporated by reference in their entirety for all purposes.

10

BACKGROUND

15 Zinc finger proteins (ZFPs) are proteins that can bind to DNA in a sequence-specific manner. Zinc fingers were first identified in the transcription factor TFIIIA from the oocytes of the African clawed toad, *Xenopus laevis*. An exemplary motif characterizing one class of these protein (C₂H₂ class) is -Cys-(X)₂₋₄-Cys-(X)₁₂-His-(X)₃₋₅-His (where X is any amino acid) (SEQ. ID. No:1). A single finger domain is about 30 amino acids in length, and several structural studies have demonstrated that it contains an alpha helix containing the two invariant histidine residues and two invariant cysteine residues in a beta turn co-ordinated through zinc. To date, over 10,000 zinc finger sequences have been identified in several thousand known or putative transcription factors. Zinc finger domains are involved not only in DNA-recognition, but also in RNA binding and in protein-protein binding. Current estimates are that this class of molecules will constitute about 2% of all human genes.

20 The x-ray crystal structure of Zif268, a three-finger domain from a murine transcription factor, has been solved in complex with a cognate DNA-sequence and shows that each finger can be superimposed on the next by a periodic rotation. The structure suggests that each finger interacts independently with DNA over 3 base-pair intervals, with side-chains at positions -1, 2, 3 and 6 on each recognition helix making contacts with their respective DNA triplet subsites. The amino terminus of Zif268 is situated at the 3' end of the DNA strand with which it makes most contacts. Some zinc fingers can bind to a fourth base in a target segment. If the strand with which a zinc finger protein makes most contacts is designated the target strand, some zinc finger

25

30

proteins bind to a three base triplet in the target strand and a fourth base on the nontarget strand. The fourth base is complementary to the base immediately 3' of the three base subsite.

The structure of the Zif268-DNA complex also suggested that the DNA sequence specificity of a zinc finger protein might be altered by making amino acid substitutions at the four helix positions (-1, 2, 3 and 6) on each of the zinc finger recognition helices. Phage display experiments using zinc finger combinatorial libraries to test this observation were published in a series of papers in 1994 (Rebar et al., *Science* 263, 671-673 (1994); Jamieson et al., *Biochemistry* 33, 5689-5695 (1994); Choo et al, *PNAS* 91, 11163-11167 (1994)). Combinatorial libraries were constructed with randomized side-chains in either the first or middle finger of Zif268 and then used to select for an altered Zif268 binding site in which the appropriate DNA sub-site was replaced by an altered DNA triplet. Further, correlation between the nature of introduced mutations and the resulting alteration in binding specificity gave rise to a partial set of substitution rules for design of ZFPs with altered binding specificity.

Greisman & Pabo, *Science* 275, 657-661 (1997) discuss an elaboration of the phage display method in which each finger of a Zif268 was successively randomized and selected for binding to a new triplet sequence. This paper reported selection of ZFPs for a nuclear hormone response element, a p53 target site and a TATA box sequence.

A number of papers have reported attempts to produce ZFPs to modulate particular target sites. For example, Choo et al., *Nature* 372, 645 (1994), report an attempt to design a ZFP that would repress expression of a bcr-abl oncogene. The target segment to which the ZFPs would bind was a nine base sequence 5'GCA GAA GCC3' chosen to overlap the junction created by a specific oncogenic translocation fusing the genes encoding bcr and abl. The intention was that a ZFP specific to this target site would bind to the oncogene without binding to abl or bcr component genes. The authors used phage display to screen a mini-library of variant ZFPs for binding to this target segment. A variant ZFP thus isolated was then reported to repress expression of a stably transfected bcr-abl construct in a cell line.

Pomerantz et al., *Science* 267, 93-96 (1995) reported an attempt to design a novel DNA binding protein by fusing two fingers from Zif268 with a homeodomain from Oct-1. The hybrid protein was then fused with a transcriptional activator for expression as a chimeric protein. The chimeric protein was reported to bind a target site

representing a hybrid of the subsites of its two components. The authors then constructed a reporter vector containing a luciferase gene operably linked to a promoter and a hybrid site for the chimeric DNA binding protein in proximity to the promoter. The authors reported that their chimeric DNA binding protein could activate expression of the luciferase gene.

Liu et al., *PNAS* 94, 5525-5530 (1997) report forming a composite zinc finger protein by using a peptide spacer to link two component zinc finger proteins each having three fingers. The composite protein was then further linked to transcriptional activation domain. It was reported that the resulting chimeric protein bound to a target site formed from the target segments bound by the two component zinc finger proteins. It was further reported that the chimeric zinc finger protein could activate transcription of a reporter gene when its target site was inserted into a reporter plasmid in proximity to a promoter operably linked to the reporter.

Choo et al., WO 98/53058, WO98/53059, and WO 98/53060 (1998) discuss selection of zinc finger proteins to bind to a target site within the HIV Tat gene. Choo et al. also discuss selection of a zinc finger protein to bind to a target site encompassing a site of a common mutation in the oncogene ras. The target site within ras was thus constrained by the position of the mutation.

The present application is related to commonly owned copending applications 09/229,007 filed January 12, 1999 and 09/229,037 filed January 12, 1999.

SUMMARY OF THE CLAIMED INVENTION

Tables 1-5 show the amino acid sequences of a large collection of zinc finger proteins and corresponding target sites bound by the proteins. Nucleotide sequences of target sites are shown in Col. 2. Target sites typically have 9 or 10 bases and constitute three target subsites bound by respective zinc finger components of a multifinger protein. Amino acid sequences of zinc finger components are shown in cols. 4, 6 and 8. The amino acids shown occupy positions -1 to +6 of a zinc finger. Table 6 shows consensus sequences for zinc fingers and target subsites bound by the fingers. Col. 1 shows the nucleotides occupying a target subsite. Cols. 2-4 show amino acids occupying positions -1 to +6 of zinc fingers binding to a target subsite.

Accordingly, the invention provides zinc fingers having amino acid sequences and target subsite binding specificities shown in Table 6. As an example, a zinc

finger having the amino acid sequence DXSNXXR at positions -1 to +6 has a target subsite GAC. As an other example, a zinc finger having the amino acid sequence RX(D/S)NXXR at positions -1 to +6 has a target subsite of GAG. A zinc finger having an amino acid sequence TXGNXXR at positions -1 to +6 has the target subsite GAT. A zinc finger having the sequence (Q/T)XSNXXR at positions -1 to +6 binds to a target subsite GAT. A zinc finger having an amino acid sequence QXG(S/D)XXR at positions -1 to +6 binds to a target subsite GCA. A zinc finger having an amino acid sequence RXDEXXR binds to a target subsite GCG. A zinc finger having an amino acid sequence QXSDXXR at positions -1 to +6 binds to a target subsite GCT. A zinc finger having an amino acid sequence QX(G/A)HXXR at positions -1 to +6 binds to a target subsite GGA. A zinc finger having an amino acid sequence DXSHXXR binds to a target subsite GGC. A zinc finger having an amino acid sequence RXDHXXR at positions -1 to +6 binds to a target subsite GGG. A zinc finger having an amino acid sequence RXDAXXR at positions -1 to +6 binds to a target subsite GTG.

The invention further provides nucleic acid encoding zinc fingers, including all of the zinc fingers described above.

The invention further provides segments of a zinc finger comprising a sequence of seven contiguous amino acids as shown in any of Tables 1-5. The invention also provides nucleic acids encoding any of these segments and zinc fingers comprising the same.

The invention further provides zinc finger proteins comprising first, second and third zinc fingers. The first, second and third zinc fingers comprise respectively first, second and third segments of seven contiguous amino acids as shown in a row of Tables 1-5. The invention further provides nucleic acids encoding such zinc finger proteins.

BRIEF DESCRIPTION OF THE FIGURE

Fig. 1 shows assembly of nucleic acids encoding zinc finger binding proteins.

DEFINITIONS

A zinc finger DNA binding protein is a protein or segment within a larger protein that binds DNA in a sequence-specific manner as a result of stabilization of protein structure through coordination of a zinc ion. The term zinc finger DNA binding protein is often abbreviated as zinc finger protein or ZFP.

A designed zinc finger protein is a protein not occurring in nature whose design/composition results principally from rational criteria. Rational criteria for design include application of substitution rules and computerized algorithms for processing information in a database storing information of existing ZFP designs and binding data. .

5 A selected zinc finger protein is a protein not found in nature whose production results primarily from an empirical process such as phage display.

The term naturally-occurring is used to describe an object that can be found in nature as distinct from being artificially produced by man. For example, a polypeptide or polynucleotide sequence that is present in an organism (including viruses) that can be isolated from a source in nature and which has not been intentionally modified by man in the laboratory is naturally-occurring. Generally, the term naturally-occurring refers to an object as present in a non-pathological (undiseased) individual, such as would be typical for the species.

15 A nucleic acid is operably linked when it is placed into a functional relationship with another nucleic acid sequence. For instance, a promoter or enhancer is operably linked to a coding sequence if it increases the transcription of the coding sequence. Operably linked means that the DNA sequences being linked are typically contiguous and, where necessary to join two protein coding regions, contiguous and in reading frame. However, since enhancers generally function when separated from the promoter by up to several kilobases or more and intronic sequences may be of variable lengths, some polynucleotide elements may be operably linked but not contiguous.

20 A specific binding affinity between, for example, a ZFP and a specific target site means a binding affinity of at least $1 \times 10^6 \text{ M}^{-1}$.

25 The terms "modulating expression" "inhibiting expression" and "activating expression" of a gene refer to the ability of a zinc finger protein to activate or inhibit transcription of a gene. Activation includes prevention of subsequent transcriptional inhibition (i.e., prevention of repression of gene expression) and inhibition includes prevention of subsequent transcriptional activation (i.e., prevention of gene activation). Modulation can be assayed by determining any parameter that is indirectly or directly affected by the expression of the target gene. Such parameters include, e.g., changes in RNA or protein levels, changes in protein activity, changes in product levels, changes in downstream gene expression, changes in reporter gene transcription (luciferase, CAT, beta-galactosidase, GFP (see, e.g., Mistili & Spector, *Nature Biotechnology* 15:961-964

(1997)); changes in signal transduction, phosphorylation and dephosphorylation, receptor-ligand interactions, second messenger concentrations (e.g., cGMP, cAMP, IP3, and Ca²⁺), cell growth, neovascularization, *in vitro*, *in vivo*, and *ex vivo*. Such functional effects can be measured by any means known to those skilled in the art, e.g.,

5 measurement of RNA or protein levels, measurement of RNA stability, identification of downstream or reporter gene expression, e.g., via chemiluminescence, fluorescence, colorimetric reactions, antibody binding, inducible markers, ligand binding assays; changes in intracellular second messengers such as cGMP and inositol triphosphate (IP3); changes in intracellular calcium levels; cytokine release, and the like.

10 A "regulatory domain" refers to a protein or a protein subsequence that has transcriptional modulation activity. Typically, a regulatory domain is covalently or non-covalently linked to a ZFP to modulate transcription. Alternatively, a ZFP can act alone, without a regulatory domain, or with multiple regulatory domains to modulate transcription.

15 A D-able subsite within a target site has the motif 5'NNGK3'. A target site containing one or more such motifs is sometimes described as a D-able target site. A zinc finger appropriately designed to bind to a D-able subsite is sometimes referred to as a D-able finger. Likewise a zinc finger protein containing at least one finger designed or selected to bind to a target site including at least one D-able subsite is sometimes referred to as a D-able zinc finger protein.

20

DETAILED DESCRIPTION

I. General

25 Tables 1-5 list a collection of nonnaturally occurring zinc finger protein sequences and their corresponding target sites. The first column of each table is an internal reference number. The second column lists a 9 or 10 base target site bound by a three-finger zinc finger protein, with the target sites listed in 5' to 3' orientation. The third column provides SEQ ID NOs for the target site sequences listed in column 2. The fourth, sixth and eighth columns list amino acid residues from the first, second and third
30 fingers, respectively, of a zinc finger protein which recognizes the target sequence listed in the second column. For each finger, seven amino acids, occupying positions -1 to +6 of the finger, are listed. The numbering convention for zinc fingers is defined below. Columns 5, 7 and 9 provide SEQ ID NOs for the amino acid sequences listed in columns

4, 6 and 8, respectively. The final column of each table lists the binding affinity (*i.e.*, the K_d in nM) of the zinc finger protein for its target site. Binding affinities are measured as described below.

Each finger binds to a triplet of bases within a corresponding target sequence. The first finger binds to the first triplet starting from the 3' end of a target site, the second finger binds to the second triplet, and the third finger binds the third (*i.e.*, the 5'-most) triplet of the target sequence. For example, the RSDSLTS finger (SEQ ID NO: 646) of SBS# 201 (Table 2) binds to 5'TTG3', the ERSTLTR finger (SEQ ID NO: 851) binds to 5'GCC3' and the QRADLRR finger (SEQ ID NO: 1056) binds to 5'GCA3'.

Table 6 lists a collection of consensus sequences for zinc fingers and the target sites bound by such sequences. Conventional one letter amino acid codes are used to designate amino acids occupying consensus positions. The symbol "X" designates a nonconsensus position that can in principle be occupied by any amino acid. In most zinc fingers of the C_2H_2 type, binding specificity is principally conferred by residues -1, +2, +3 and +6. Accordingly, consensus sequence determining binding specificity typically include at least these residues. Consensus sequences are useful for designing zinc fingers to bind to a given target sequence. Residues occupying other positions can be selected based on sequences in Tables 1-5, or other known zinc finger sequences. Alternatively, these positions can be randomized with a plurality of candidate amino acids and screened against one or more target sequences to refine binding specificity or improve binding specificity. In general, the same consensus sequence can be used for design of a zinc finger regardless of the relative position of that finger in a multi-finger zinc finger protein. For example, the sequence RXDNXXR can be used to design a N-terminal, central or C-terminal finger of three finger protein. However, some consensus sequences are most suitable for designing a zinc finger to occupy a particular position in a multi-finger protein. For example, the consensus sequence RXDHXXQ is most suitable for designing a C-terminal finger of a three-finger protein.

II. Characteristics of Zinc Finger Proteins

Zinc finger proteins are formed from zinc finger components. For example, zinc finger proteins can have one to thirty-seven fingers, commonly having 2, 3, 4, 5 or 6 fingers. A zinc finger protein recognizes and binds to a target site (sometimes

referred to as a target segment) that represents a relatively small subsequence within a target gene. Each component finger of a zinc finger protein can bind to a subsite within the target site. The subsite includes a triplet of three contiguous bases all on the same strand (sometimes referred to as the target strand). The subsite may or may not also include a fourth base on the opposite strand that is the complement of the base immediately 3' of the three contiguous bases on the target strand. In many zinc finger proteins, a zinc finger binds to its triplet subsite substantially independently of other fingers in the same zinc finger protein. Accordingly, the binding specificity of zinc finger protein containing multiple fingers is usually approximately the aggregate of the specificities of its component fingers. For example, if a zinc finger protein is formed from first, second and third fingers that individually bind to triplets XXX, YYY, and ZZZ, the binding specificity of the zinc finger protein is 3'XXX YYY ZZZ5'.

The relative order of fingers in a zinc finger protein from N-terminal to C-terminal determines the relative order of triplets in the 3' to 5' direction in the target. For example, if a zinc finger protein comprises from N-terminal to C-terminal first, second and third fingers that individually bind, respectively, to triplets 5' GAC3', 5'GTA3' and 5'GGC3' then the zinc finger protein binds to the target segment 3'CAGATGCGG5'. If the zinc finger protein comprises the fingers in another order, for example, second finger, first finger, third finger, then the zinc finger protein binds to a target segment comprising a different permutation of triplets, in this example, 3'ATGCAGCGG5' (see Berg & Shi, *Science* 271, 1081-1086 (1996)). The assessment of binding properties of a zinc finger protein as the aggregate of its component fingers may, in some cases, be influenced by context-dependent interactions of multiple fingers binding in the same protein.

Two or more zinc finger proteins can be linked to have a target specificity that is the aggregate of that of the component zinc finger proteins (see e.g., Kim & Pabo, *PNAS* 95, 2812-2817 (1998)). For example, a first zinc finger protein having first, second and third component fingers that respectively bind to XXX, YYY and ZZZ can be linked to a second zinc finger protein having first, second and third component fingers with binding specificities, AAA, BBB and CCC. The binding specificity of the combined first and second proteins is thus 3'XXXYYYZZZ___AAABBBCCC5', where the underline indicates a short intervening region (typically 0-5 bases of any type). In this situation, the target site can be viewed as comprising two target segments separated by an intervening segment.

Linkage can be accomplished using any of the following peptide linkers.

T G E K P: (SEQ. ID. No:2) (Liu et al., 1997, supra.); (G4S)_n (SEQ. ID. No:3) (Kim et al., *PNAS* 93, 1156-1160 (1996.); GGRRGGGS; (SEQ. ID. No:4) LRQRDGERP; (SEQ. ID. No:5) LRQKDGGGSERP; (SEQ. ID. No:6) LRQKD(G3S)₂ ERP (SEQ. ID. No:7)

5 Alternatively, flexible linkers can be rationally designed using computer programs capable of modeling both DNA-binding sites and the peptides themselves or by phage display methods . In a further variation, noncovalent linkage can be achieved by fusing two zinc finger proteins with domains promoting heterodimer formation of the two zinc finger proteins. For example, one zinc finger protein can be fused with fos and the other
10 with jun (see Barbas et al., WO 95/119431).

Linkage of two zinc finger proteins is advantageous for conferring a unique binding specificity within a mammalian genome. A typical mammalian diploid genome consists of 3×10^9 bp. Assuming that the four nucleotides A, C, G, and T are randomly distributed, a given 9 bp sequence is present ~23,000 times. Thus a ZFP
15 recognizing a 9 bp target with absolute specificity would have the potential to bind to ~23,000 sites within the genome. An 18 bp sequence is present once in 3.4×10^{10} bp, or about once in a random DNA sequence whose complexity is ten times that of a mammalian genome.

A component finger of zinc finger protein typically contains about 30
20 amino acids and has the following motif (N-C) :

(SEQ. ID. No:8)

Cys- (X)₂₋₄-Cys-X.X.X.X.X.X.X.X.X.X.X.X.X-**His**- (X)₃₋₅-His
-1 1 2 3 4 5 6 7

The two invariant histidine residues and two invariant cysteine residues in
25 a single beta turn are co-ordinated through zinc (see, e.g., Berg & Shi, *Science* 271, 1081-1085 (1996)). The above motif shows a numbering convention that is standard in the field for the region of a zinc finger conferring binding specificity. The amino acid on the left (N-terminal side) of the first invariant His residues is assigned the number +6, and other amino acids further to the left are assigned successively decreasing numbers. The
30 alpha helix begins at residue 1 and extends to the residue following the second conserved histidine. The entire helix is therefore of variable length, between 11 and 13 residues.

The process of designing or selecting a nonnaturally occurring or variant ZFP typically starts with a natural ZFP as a source of framework residues. The process of

design or selection serves to define nonconserved positions (i.e., positions -1 to +6) so as to confer a desired binding specificity. One suitable ZFP is the DNA binding domain of the mouse transcription factor Zif268. The DNA binding domain of this protein has the amino acid sequence:

- 5 YACPVESCDRRFSRSDDELTRHIRIHTGQKP (F1) (SEQ. ID No:9)
 FQCRICMRNFSRSDHLTTHIRTHHTGEKP (F2) (SEQ. ID. No:10)
 FACDICGRKFARSDERKRHTKIHLRQK (F3) SEQ. ID. No:11)
 and binds to a target 5' GCG TGG GCG 3' (SEQ ID No:12).

- Another suitable natural zinc finger protein as a source of framework
 10 residues is Sp-1. The Sp-1 sequence used for construction of zinc finger proteins corresponds to amino acids 531 to 624 in the Sp-1 transcription factor. This sequence is 94 amino acids in length. The amino acid sequence of Sp-1 is as follows:

PGKKKQHICHIQGCGKVYGKTSHLRAHLRWHTGERP
 FMCTWSYCGKRFTSRDELQRHKRTHHTGEKK

- 15 FACPECPKRFMRSDHLSKHIKTHQNKKG (SEQ. ID. No:13)
 Sp-1 binds to a target site 5'GGG GCG GGG3' (SEQ ID No: 14).

An alternate form of Sp-1, an Sp-1 consensus sequence, has the following amino acid sequence:

meklrngsgd

- 20 PGKKKQHACPECCKSFSKSSHLRAHQRTHTGERP
 YKCPECCKSFSRSDHLQRHQRTHTGEKP
 YKCPECCKSFSRSDHLSKHQRTHTQNKKG (SEQ. ID. No:15) (lower case letters are a leader sequence from Shi & Berg, *Chemistry and Biology* 1, 83-89. (1995). The optimal binding sequence for the Sp-1 consensus sequence is 5'GGGGCGGGG3' (SEQ ID No: 16) . Other suitable ZFPs are described below.

- 25 There are a number of substitution rules that assist rational design of some zinc finger proteins (see Desjarlais & Berg, *PNAS* 90, 2256-2260 (1993); Choo & Klug, *PNAS* 91, 11163-11167 (1994); Desjarlais & Berg, *PNAS* 89, 7345-7349 (1992); Jamieson et al., supra; Choo et al., WO 98/53057, WO 98/53058; WO 98/53059; WO
 30 98/53060). Many of these rules are supported by site-directed mutagenesis of the three-finger domain of the ubiquitous transcription factor, Sp-1 (Desjarlais and Berg, 1992; 1993). One of these rules is that a 5' G in a DNA triplet can be bound by a zinc finger incorporating arginine at position 6 of the recognition helix. Another substitution rule is

that a G in the middle of a subsite can be recognized by including a histidine residue at position 3 of a zinc finger. A further substitution rule is that asparagine can be incorporated to recognize A in the middle of triplet, aspartic acid, glutamic acid, serine or threonine can be incorporated to recognize C in the middle of triplet, and amino acids with small side chains such as alanine can be incorporated to recognize T in the middle of triplet. A further substitution rule is that the 3' base of triplet subsite can be recognized by incorporating the following amino acids at position -1 of the recognition helix: arginine to recognize G, glutamine to recognize A, glutamic acid (or aspartic acid) to recognize C, and threonine to recognize T. Although these substitution rules are useful in designing zinc finger proteins they do not take into account all possible target sites. Furthermore, the assumption underlying the rules, namely that a particular amino acid in a zinc finger is responsible for binding to a particular base in a subsite is only approximate. Context-dependent interactions between proximate amino acids in a finger or binding of multiple amino acids to a single base or vice versa can cause variation of the binding specificities predicted by the existing substitution rules.

The technique of phage display provides a largely empirical means of generating zinc finger proteins with a desired target specificity (see e.g., Rebar, US 5,789,538; Choo et al., WO 96/06166; Barbas et al., WO 95/19431 and WO 98/543111; Jamieson et al., supra). The method can be used in conjunction with, or as an alternative to rational design. The method involves the generation of diverse libraries of mutagenized zinc finger proteins, followed by the isolation of proteins with desired DNA-binding properties using affinity selection methods. To use this method, the experimenter typically proceeds as follows. First, a gene for a zinc finger protein is mutagenized to introduce diversity into regions important for binding specificity and/or affinity. In a typical application, this is accomplished via randomization of a single finger at positions -1, +2, +3, and +6, and sometimes accessory positions such as +1, +5, +8 and +10. Next, the mutagenized gene is cloned into a phage or phagemid vector as a fusion with gene III of a filamentous phage, which encodes the coat protein pIII. The zinc finger gene is inserted between segments of gene III encoding the membrane export signal peptide and the remainder of pIII, so that the zinc finger protein is expressed as an amino-terminal fusion with pIII or in the mature, processed protein. When using phagemid vectors, the mutagenized zinc finger gene may also be fused to a truncated version of gene III encoding, minimally, the C-terminal region required for assembly of pIII into the phage

particle. The resultant vector library is transformed into *E. coli* and used to produce filamentous phage which express variant zinc finger proteins on their surface as fusions with the coat protein pIII. If a phagemid vector is used, then this step requires superinfection with helper phage. The phage library is then incubated with target DNA site, and affinity selection methods are used to isolate phage which bind target with high affinity from bulk phage. Typically, the DNA target is immobilized on a solid support, which is then washed under conditions sufficient to remove all but the tightest binding phage. After washing, any phage remaining on the support are recovered via elution under conditions which disrupt zinc finger – DNA binding. Recovered phage are used to infect fresh *E. coli*, which is then amplified and used to produce a new batch of phage particles. Selection and amplification are then repeated as many times as is necessary to enrich the phage pool for tight binders such that these may be identified using sequencing and/or screening methods. Although the method is illustrated for pIII fusions, analogous principles can be used to screen ZFP variants as pVIII fusions.

In certain embodiments, the sequence bound by a particular zinc finger protein is determined by conducting binding reactions (see, *e.g.*, conditions for determination of K_d , *infra*) between the protein and a pool of randomized double-stranded oligonucleotide sequences. The binding reaction is analyzed by an electrophoretic mobility shift assay (EMSA), in which protein-DNA complexes undergo retarded migration in a gel and can be separated from unbound nucleic acid. Oligonucleotides which have bound the finger are purified from the gel and amplified, for example, by a polymerase chain reaction. The selection (*i.e.* binding reaction and EMSA analysis) is then repeated as many times as desired, with the selected oligonucleotide sequences. In this way, the binding specificity of a zinc finger protein having a particular amino acid sequence is determined.

Zinc finger proteins are often expressed with a heterologous domain as fusion proteins. Common domains for addition to the ZFP include, *e.g.*, transcription factor domains (activators, repressors, co-activators, co-repressors), silencers, oncogenes (*e.g.*, myc, jun, fos, myb, max, mad, rel, ets, bcl, myb, mos family members etc.); DNA repair enzymes and their associated factors and modifiers; DNA rearrangement enzymes and their associated factors and modifiers; chromatin associated proteins and their modifiers (*e.g.* kinases, acetylases and deacetylases); and DNA modifying enzymes (*e.g.*, methyltransferases, topoisomerases, helicases, ligases, kinases, phosphatases,

polymerases, endonucleases) and their associated factors and modifiers. A preferred domain for fusing with a ZFP when the ZFP is to be used for repressing expression of a target gene is a KRAB repression domain from the human KOX-1 protein (Thiesen et al., *New Biologist* 2, 363-374 (1990); Margolin et al., *Proc. Natl. Acad. Sci. USA* 91, 4509-4513 (1994); Pengue et al., *Nucl. Acids Res.* 22:2908-2914 (1994); Witzgall et al., *Proc. Natl. Acad. Sci. USA* 91, 4514-4518 (1994). Preferred domains for achieving activation include the HSV VP16 activation domain (see, e.g., Hagmann et al., *J. Virol.* 71, 5952-5962 (1997)) nuclear hormone receptors (see, e.g., Torchia et al., *Curr. Opin. Cell. Biol.* 10:373-383 (1998)); the p65 subunit of nuclear factor kappa B (Bitko & Barik, *J. Virol.* 72:5610-5618 (1998) and Doyle & Hunt, *Neuroreport* 8:2937-2942 (1997)); Liu et al., *Cancer Gene Ther.* 5:3-28 (1998)), or artificial chimeric functional domains such as VP64 (Seifpal et al., *EMBO J.* 11, 4961-4968 (1992)).

An important factor in the administration of polypeptide compounds, such as the ZFPs, is ensuring that the polypeptide has the ability to traverse the plasma membrane of a cell, or the membrane of an intra-cellular compartment such as the nucleus. Cellular membranes are composed of lipid-protein bilayers that are freely permeable to small, nonionic lipophilic compounds and are inherently impermeable to polar compounds, macromolecules, and therapeutic or diagnostic agents. However, proteins and other compounds such as liposomes have been described, which have the ability to translocate polypeptides such as ZFPs across a cell membrane.

For example, "membrane translocation polypeptides" have amphiphilic or hydrophobic amino acid subsequences that have the ability to act as membrane-translocating carriers. In one embodiment, homeodomain proteins have the ability to translocate across cell membranes. The shortest internalizable peptide of a homeodomain protein, Antennapedia, was found to be the third helix of the protein, from amino acid position 43 to 58 (see, e.g., Prochiantz, *Current Opinion in Neurobiology* 6:629-634 (1996)). Another subsequence, the h (hydrophobic) domain of signal peptides, was found to have similar cell membrane translocation characteristics (see, e.g., Lin et al., *J. Biol. Chem.* 270:14255-14258 (1995)).

Examples of peptide sequences which can be linked to a ZFP of the invention, for facilitating uptake of ZFP into cells, include, but are not limited to: an 11 amino acid peptide of the tat protein of HIV; a 20 residue peptide sequence which corresponds to amino acids 84-103 of the p16 protein (see Fahraeus et al., *Current*

Biology 6:84 (1996)); the third helix of the 60-amino acid long homeodomain of Antennapedia (Derossi *et al.*, *J. Biol. Chem.* 269:10444 (1994)); the h region of a signal peptide such as the Kaposi fibroblast growth factor (K-FGF) h region (Lin *et al.*, *supra*); or the VP22 translocation domain from HSV (Elliot & O'Hare, *Cell* 88:223-233 (1997)).

5 Other suitable chemical moieties that provide enhanced cellular uptake may also be chemically linked to ZFPs.

Toxin molecules also have the ability to transport polypeptides across cell membranes. Often, such molecules are composed of at least two parts (called "binary toxins"): a translocation or binding domain or polypeptide and a separate toxin domain or polypeptide. Typically, the translocation domain or polypeptide binds to a cellular
10 receptor, and then the toxin is transported into the cell. Several bacterial toxins, including *Clostridium perfringens* iota toxin, diphtheria toxin (DT), *Pseudomonas* exotoxin A (PE), pertussis toxin (PT), *Bacillus anthracis* toxin, and pertussis adenylate cyclase (CYA), have been used in attempts to deliver peptides to the cell cytosol as internal or amino-
15 terminal fusions (Arora *et al.*, *J. Biol. Chem.*, 268:3334-3341 (1993); Perelle *et al.*, *Infect. Immun.*, 61:5147-5156 (1993); Stenmark *et al.*, *J. Cell Biol.* 113:1025-1032 (1991); Donnelly *et al.*, *PNAS* 90:3530-3534 (1993); Carbonetti *et al.*, *Abstr. Annu. Meet. Am. Soc. Microbiol.* 95:295 (1995); Sebo *et al.*, *Infect. Immun.* 63:3851-3857 (1995); Klimpel *et al.*, *PNAS U.S.A.* 89:10277-10281 (1992); and Novak *et al.*, *J. Biol. Chem.* 267:17186-
20 17193 1992)).

Such subsequences can be used to translocate ZFPs across a cell membrane. ZFPs can be conveniently fused to or derivatized with such sequences. Typically, the translocation sequence is provided as part of a fusion protein. Optionally, a linker can be used to link the ZFP and the translocation sequence. Any suitable linker can
25 be used, e.g., a peptide linker.

Production of ZFPs

ZFP polypeptides and nucleic acids encoding the same can be made using routine techniques in the field of recombinant genetics. Basic texts disclosing the general
30 methods of use in this invention include Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2nd ed. 1989); Kriegler, *Gene Transfer and Expression: A Laboratory Manual* (1990); and *Current Protocols in Molecular Biology* (Ausubel *et al.*, eds., 1994)). In addition, nucleic acids less than about 100 bases can be custom ordered

from any of a variety of commercial sources, such as The Midland Certified Reagent Company (mcrc@oligos.com), The Great American Gene Company (<http://www.genco.com>), ExpressGen Inc. (www.expressgen.com), Operon Technologies Inc. (Alameda, CA). Similarly, peptides can be custom ordered from any of a variety of sources, such as PeptidoGenic (pkim@ccnet.com), HTI Bio-products, inc. (<http://www.htibio.com>), BMA Biomedicals Ltd (U.K.), Bio.Synthesis, Inc.

Oligonucleotides can be chemically synthesized according to the solid phase phosphoramidite triester method first described by Beaucage & Caruthers, *Tetrahedron Letts.* 22:1859-1862 (1981), using an automated synthesizer, as described in Van Devanter et al., *Nucleic Acids Res.* 12:6159-6168 (1984). Purification of oligonucleotides is by either denaturing polyacrylamide gel electrophoresis or by reverse phase HPLC. The sequence of the cloned genes and synthetic oligonucleotides can be verified after cloning using, e.g., the chain termination method for sequencing double-stranded templates of Wallace et al., *Gene* 16:21-26 (1981).

Two alternative methods are typically used to create the coding sequences required to express newly designed DNA-binding peptides. One protocol is a PCR-based assembly procedure that utilizes six overlapping oligonucleotides (Fig. 1). Three oligonucleotides (oligos 1, 3, and 5 in Figure 1) correspond to “universal” sequences that encode portions of the DNA-binding domain between the recognition helices. These oligonucleotides typically remain constant for all zinc finger constructs. The other three “specific” oligonucleotides (oligos 2, 4, and 6 in Fig. 1) are designed to encode the recognition helices. These oligonucleotides contain substitutions primarily at positions - 1, 2, 3 and 6 on the recognition helices making them specific for each of the different DNA-binding domains.

The PCR synthesis is carried out in two steps. First, a double stranded DNA template is created by combining the six oligonucleotides (three universal, three specific) in a four cycle PCR reaction with a low temperature annealing step, thereby annealing the oligonucleotides to form a DNA “scaffold.” The gaps in the scaffold are filled in by high-fidelity thermostable polymerase, the combination of Taq and Pfu polymerases also suffices. In the second phase of construction, the zinc finger template is amplified by external primers designed to incorporate restriction sites at either end for cloning into a shuttle vector or directly into an expression vector.

An alternative method of cloning the newly designed DNA-binding proteins relies on annealing complementary oligonucleotides encoding the specific regions of the desired ZFP. This particular application requires that the oligonucleotides be phosphorylated prior to the final ligation step. This is usually performed before setting up the annealing reactions. In brief, the “universal” oligonucleotides encoding the constant regions of the proteins (oligos 1, 2 and 3 of above) are annealed with their complementary oligonucleotides. Additionally, the “specific” oligonucleotides encoding the finger recognition helices are annealed with their respective complementary oligonucleotides. These complementary oligos are designed to fill in the region which was previously filled in by polymerase in the above-mentioned protocol. The complementary oligos to the common oligos 1 and finger 3 are engineered to leave overhanging sequences specific for the restriction sites used in cloning into the vector of choice in the following step. The second assembly protocol differs from the initial protocol in the following aspects: the “scaffold” encoding the newly designed ZFP is composed entirely of synthetic DNA thereby eliminating the polymerase fill-in step, additionally the fragment to be cloned into the vector does not require amplification. Lastly, the design of leaving sequence-specific overhangs eliminates the need for restriction enzyme digests of the inserting fragment. Alternatively, changes to ZFP recognition helices can be created using conventional site-directed mutagenesis methods.

Both assembly methods require that the resulting fragment encoding the newly designed ZFP be ligated into a vector. Ultimately, the ZFP-encoding sequence is cloned into an expression vector. Expression vectors that are commonly utilized include, but are not limited to, a modified pMAL-c2 bacterial expression vector (New England BioLabs or an eukaryotic expression vector, pcDNA (Promega). The final constructs are verified by sequence analysis.

Any suitable method of protein purification known to those of skill in the art can be used to purify ZFPs of the invention (see, Ausubel, supra, Sambrook, supra). In addition, any suitable host can be used for expression, e.g., bacterial cells, insect cells, yeast cells, mammalian cells, and the like.

Expression of a zinc finger protein fused to a maltose binding protein (MBP-ZFP) in bacterial strain JM109 allows for straightforward purification through an amylose column (NEB). High expression levels of the zinc finger chimeric protein can be obtained by induction with IPTG since the MBP-ZFP fusion in the pMal-c2 expression

plasmid is under the control of the tac promoter (NEB). Bacteria containing the MBP-ZFP fusion plasmids are inoculated into 2xYT medium containing 10 μ M ZnCl₂, 0.02% glucose, plus 50 μ g/ml ampicillin and shaken at 37°C. At mid-exponential growth IPTG is added to 0.3 mM and the cultures are allowed to shake. After 3 hours the bacteria are harvested by centrifugation, disrupted by sonication or by passage through a french pressure cell or through the use of lysozyme, and insoluble material is removed by centrifugation. The MBP-ZFP proteins are captured on an amylose-bound resin, washed extensively with buffer containing 20 mM Tris-HCl (pH 7.5), 200 mM NaCl, 5 mM DTT and 50 μ M ZnCl₂, then eluted with maltose in essentially the same buffer (purification is based on a standard protocol from NEB). Purified proteins are quantitated and stored for biochemical analysis.

The dissociation constants of the purified proteins, e.g., K_d, are typically characterized via electrophoretic mobility shift assays (EMSA) (Buratowski & Chodosh, in *Current Protocols in Molecular Biology* pp. 12.2.1-12.2.7 (Ausubel ed., 1996)).

Affinity is measured by titrating purified protein against a fixed amount of labeled double-stranded oligonucleotide target. The target typically comprises the natural binding site sequence flanked by the 3 bp found in the natural sequence and additional, constant flanking sequences. The natural binding site is typically 9 bp for a three-finger protein and 2 x 9 bp + intervening bases for a six finger ZFP. The annealed oligonucleotide targets possess a 1 base 5' overhang which allows for efficient labeling of the target with T4 phage polynucleotide kinase. For the assay the target is added at a concentration of 1 nM or lower (the actual concentration is kept at least 10-fold lower than the expected dissociation constant), purified ZFPs are added at various concentrations, and the reaction is allowed to equilibrate for at least 45 min. In addition the reaction mixture also contains 10 mM Tris (pH 7.5), 100 mM KCl, 1 mM MgCl₂, 0.1 mM ZnCl₂, 5 mM DTT, 10% glycerol, 0.02% BSA. (NB: in earlier assays poly d(IC) was also added at 10-100 μ g/ μ l.)

The equilibrated reactions are loaded onto a 10% polyacrylamide gel, which has been pre-run for 45 min in Tris/glycine buffer, then bound and unbound labeled target is resolved by electrophoresis at 150V. (alternatively, 10-20% gradient Tris-HCl gels, containing a 4% polyacrylamide stacker, can be used) The dried gels are visualized by autoradiography or phosphorimaging and the apparent K_d is determined by calculating the protein concentration that gives half-maximal binding.

The assays can also include determining active fractions in the protein preparations. Active fractions are determined by stoichiometric gel shifts where proteins are titrated against a high concentration of target DNA. Titrations are done at 100, 50, and 25% of target (usually at micromolar levels).

5

III. Applications of Designed ZFPs

ZFPs that bind to a particular target gene, and the nucleic acids encoding them, can be used for a variety of applications. These applications include therapeutic methods in which a ZFP or a nucleic acid encoding it is administered to a subject and used to modulate the expression of a target gene within the subject (see copending application Townsend & Townsend & Crew Attorney Docket 019496-002200, filed January 12, 1999). The modulation can be in the form of repression, for example, when the target gene resides in a pathological infecting microorganisms, or in an endogenous gene of the patient, such as an oncogene or viral receptor, that is contributing to a disease state. Alternatively, the modulation can be in the form of activation when activation of expression or increased expression of an endogenous cellular gene can ameliorate a diseased state. For such applications, ZFPs, or more typically, nucleic acids encoding them are formulated with a pharmaceutically acceptable carrier as a pharmaceutical composition.

Pharmaceutically acceptable carriers are determined in part by the particular composition being administered, as well as by the particular method used to administer the composition. (see, e.g., *Remington's Pharmaceutical Sciences*, 17th ed. 1985)). The ZFPs, alone or in combination with other suitable components, can be made into aerosol formulations (i.e., they can be "nebulized") to be administered via inhalation. Aerosol formulations can be placed into pressurized acceptable propellants, such as dichlorodifluoromethane, propane, nitrogen, and the like. Formulations suitable for parenteral administration, such as, for example, by intravenous, intramuscular, intradermal, and subcutaneous routes, include aqueous and non-aqueous, isotonic sterile injection solutions, which can contain antioxidants, buffers, bacteriostats, and solutes that render the formulation isotonic with the blood of the intended recipient, and aqueous and non-aqueous sterile suspensions that can include suspending agents, solubilizers, thickening agents, stabilizers, and preservatives. Compositions can be administered, for example, by intravenous infusion, orally, topically, intraperitoneally, intravesically or

intrathecally. The formulations of compounds can be presented in unit-dose or multi-dose sealed containers, such as ampules and vials. Injection solutions and suspensions can be prepared from sterile powders, granules, and tablets of the kind previously described.

5 The dose administered to a patient should be sufficient to effect a beneficial therapeutic response in the patient over time. The dose is determined by the efficacy and K_d of the particular ZFP employed, the target cell, and the condition of the patient, as well as the body weight or surface area of the patient to be treated. The size of the dose also is determined by the existence, nature, and extent of any adverse side-effects
10 that accompany the administration of a particular compound or vector in a particular patient

In other applications, ZFPs are used in diagnostic methods for sequence specific detection of target nucleic acid in a sample. For example, ZFPs can be used to detect variant alleles associated with a disease or phenotype in patient samples. As an
15 example, ZFPs can be used to detect the presence of particular mRNA species or cDNA in a complex mixtures of mRNAs or cDNAs. As a further example, ZFPs can be used to quantify copy number of a gene in a sample. For example, detection of loss of one copy of a p53 gene in a clinical sample is an indicator of susceptibility to cancer. In a further example, ZFPs are used to detect the presence of pathological microorganisms in clinical
20 samples. This is achieved by using one or more ZFPs specific to genes within the microorganism to be detected. A suitable format for performing diagnostic assays employs ZFPs linked to a domain that allows immobilization of the ZFP on an ELISA plate. The immobilized ZFP is contacted with a sample suspected of containing a target nucleic acid under conditions in which binding can occur. Typically, nucleic acids in the
25 sample are labeled (e.g., in the course of PCR amplification). Alternatively, unlabelled probes can be detected using a second labelled probe. After washing, bound-labelled nucleic acids are detected.

ZFPs also can be used for assays to determine the phenotype and function of gene expression. Current methodologies for determination of gene function rely
30 primarily upon either overexpression or removing (knocking out completely) the gene of interest from its natural biological setting and observing the effects. The phenotypic effects observed indicate the role of the gene in the biological system.

One advantage of ZFP-mediated regulation of a gene relative to conventional knockout analysis is that expression of the ZFP can be placed under small molecule control. By controlling expression levels of the ZFPs, one can in turn control the expression levels of a gene regulated by the ZFP to determine what degree of repression or stimulation of expression is required to achieve a given phenotypic or biochemical effect. This approach has particular value for drug development. By putting the ZFP under small molecule control, problems of embryonic lethality and developmental compensation can be avoided by switching on the ZFP repressor at a later stage in mouse development and observing the effects in the adult animal. Transgenic mice having target genes regulated by a ZFP can be produced by integration of the nucleic acid encoding the ZFP at any site *in trans* to the target gene. Accordingly, homologous recombination is not required for integration of the nucleic acid. Further, because the ZFP is trans-dominant, only one chromosomal copy is needed and therefore functional knock-out animals can be produced without backcrossing.

All references cited above are hereby incorporated by reference in their entirety for all purposes.

TABLE 1

<u>SBS#</u>	<u>TARGET</u>	SEQ ID	<u>F1</u>	SEQ ID	<u>F2</u>	SEQ ID	<u>F3</u>	SEQ ID	<u>Kd</u> (nM)
249	GCGGGGGCG	17	RSDELTR	123	RSDHLR	229	RSDELRR	335	20
250	GCGGGGGCG	18	RSDELTR	124	RSDHLR	230	RSDTLKK	336	70
251	GCGGAGGCG	19	RSDELTR	125	RSDNLTR	231	RSDELRR	337	27.5
252	GCGGCCGCG	20	RSDELTR	126	DRSSLTR	232	RSDELRR	338	100
253	GGATGGGGG	21	RSDHLAR	127	RSDHLTT	233	QRAHLAR	339	0.75
256	GCGGGGTCC	22	ERGDLT	128	RSDHLR	234	RSDELRR	340	800
258	GCGGGCGGG	23	RSDHLTR	129	ERGHLTR	235	RSDELRR	341	15
259	GCAGAGGAG	24	RSDNLAR	130	RSDNLAR	236	QSGSLTR	342	250
261	GAGGTGGCC	25	ERGTLAR	131	RSDALSR	237	RSDNLSR	343	0.5
262	GCGGGGGCT	26	QSSDLQR	132	RSDHLR	238	RSDELRR	344	20
263	GCGGGGGCT	27	QSSDLQR	133	RSDHLR	239	RSDTLKK	345	1
264	GTGGCTGCC	28	DRSSLTR	134	QSSDLQR	240	RSDALAR	346	27
265	GTGGCTGCC	29	ERGTLAR	135	QSSDLQR	241	RSDALAR	347	600
269	GGGGCCGGG	30	RSDHLTR	136	DRSSLTR	242	RSDHLTR	348	5
270	GGGGCCGGG	31	RSDHLTR	137	ERGTLAR	243	RSDHLTR	349	52.5
272	GCAGGGGCC	32	DRSSLTR	138	RSDHLR	244	QSGSLTR	350	20
337	TGCGGGGCAA	33	RSADLTR	139	RSDHLTR	245	ERQHLAT	351	24
338	TGCGGGGCAA	34	RSADLTR	140	RSDHLTR	246	ERDHLRT	352	8
339	TGCGGGGCAA	35	RSADLTR	141	RSDHLTT	247	ERQHLAT	353	64
340	TGCGGGGCAA	36	RSADLTR	142	RSDHLTT	248	ERDHLRT	354	48
341	TGCGGGGCAA	37	RSADLTR	143	RGDHLKD	249	ERQHLAT	355	1000
342	TGCGGGGCAA	38	RSADLTR	144	RGDHLKD	250	ERDHLRT	356	1000
343	TGCGGGGCAA	39	QSGSLTR	145	RSDHLTR	251	ERQHLAT	357	8
344	TGCGGGGCAA	40	QSGSLTR	146	RSDHLTR	252	ERDHLRT	358	6
345	TGCGGGGCAA	41	QSGSLTR	147	RSDHLTT	253	ERQHLAT	359	96
346	TGCGGGGCAA	42	QSGSLTR	148	RSDHLTT	254	ERDHLRT	360	64
347	TGCGGGGCAA	43	QSGSLTR	149	RGDHLKD	255	ERQHLAT	361	1000

348	TGCGGGGCAA	44	QSGSLTR	150	RGDHLKD	256	ERDHLRT	362	1000
367	GGGGGCGGG	45	RSDHLTR	151	DSGHLTR	257	RSDHLQR	363	60
368	GAGGGGGCG	46	RSDELTR	152	RSDHLTR	258	RSDNLTR	364	3.5
369	GTAGTTGTG	47	RSDALTR	153	TGGSLAR	259	QSGSLTR	365	95
370	GTAGTTGTG	48	RSDALTR	154	NRATLAR	260	QSASLTR	366	300
371	GTAGTTGTG	49	RSDALTR	155	NRATLAR	261	QSGSLTR	367	175
372	GTAGTTGTG	50	RSDSLLR	156	TGGSLAR	262	QSASLTR	368	112.5
373	GTAGTTGTG	51	RSDSLLR	157	NRATLAR	263	QSASLTR	369	320
374	GCTGAGGAA	52	QRSNLVR	158	RSDNLTR	264	TSSELQR	370	3.3
375	GAGGAAGAT	53	QQSNLAR	159	QSGNLQR	265	RSDNLTR	371	85
401	GTAGTTGTG	54	RSDALTR	160	TGGSLAR	266	QSASLTR	372	80
403	GTAGTTGTG	55	RSDSLLR	161	NRATLAR	267	QSGSLTR	373	750
421	GTAGTTGTG	56	DSDSLLR	162	TGGSLAR	268	QSGSLTR	374	500
422	GTAGTTGTG	57	RSDSLLR	163	TGGSLTR	269	QSGSLTR	375	200
423	GTAGTTGTG	58	RSDALTR	164	TGGSLAR	270	QRSALAR	376	1000
424	GATGCTGAG	59	RSDNLTR	165	TSSELQR	271	TSANLSR	377	100
425	GATGCTGAG	60	RSDNLTR	166	QSSDLQR	272	QQSNLAR	378	25
426	GATGCTGAG	61	RSDNLTR	167	QSSDLQR	273	TSANLSR	379	5.5
427	GCTGAGGAA	62	QRSNLVR	168	RSDNLTR	274	QSSDLQR	380	1
428	GAAGATGAC	63	DSSNLTR	169	QQSNLAR	275	QRSNLVR	381	120
429	GAAGATGAC	64	DSSNLTR	170	TSANLSR	276	QRSNLVR	382	50
430	GATGACGAC	65	EKANLTR	171	DSSNLTR	277	QQSNLAR	383	250
431	GACGACGGC	66	DSGHLTR	172	DRSNLER	278	DSSNLTR	384	100
432	GACGACGGC	67	DSGHLTR	173	DHANLAR	279	DSSNLTR	385	1000
433	GACGACGGC	68	DSGNLTR	174	DHANLAR	280	DSSNLTR	386	1000
434	GACGGCGTA	69	QSASLTR	175	DSGHLTR	281	EKANLTR	387	152.5
435	GACGGCGTA	70	QSASLTR	176	DSGHLTR	282	ERGNLTR	388	150
436	GACGGCGTA	71	QRSALAR	177	DSGHLTR	283	EKANLTR	389	95
437	GACGGCGTA	72	QRSALAR	178	DSGHLTR	284	ERGNLTR	390	117.5
438	GAGGGGGCG	73	RSDELTR	179	RSDHLTT	285	RSDNLTR	391	62.5
440	GCCGAGGTGC	74	RSDSLLR	180	RSKNLQR	286	ERGTLAR	392	40
441	GGTGGAGTCA	75	DSGSLTR	181	QSGHLQR	287	TSGHLTR	393	250
445	GTCGCACTGA	76	RSDSLRR	182	QSSDLQK	288	DSGSLTR	394	1000

450	GACTTGGTGC	77	RSDTLAR	183	RGDALTS	289	DRSNLTR	395	130
453	GGTGGAGTCA	78	DRSALAR	184	QSGHLQR	290	DSSKLSR	396	150
461	GAGTACTGTA	79	QRSHLTT	185	DRSNLRT	291	RSDNLAR	397	120
463	GTGGAGGAGA	80	RSDNLTR	186	RSDNLAR	292	RSDALAR	398	0.5
464	GTGGAGGAGA	81	RSDNLTR	187	RSDNLAR	293	RSDSLAR	399	0.4
466	CAGGCTGCGC	82	RSDDLTR	188	QSSDLQR	294	RSDNLRE	400	65
467	CAGGCTGCGC	83	RSDELTR	189	QSSDLQR	295	RGDHLKD	401	800
468	CAGGCTGCGC	84	RSDDLTR	190	QSSDLQR	296	RGDHLKD	402	42
469	GAAGAGGTCT	85	DRSALAR	191	RSDNLAR	297	QSGNLTR	403	13.5
472	GAGGTCTGGA	86	RSSHITT	192	DRSALAR	298	RSDNLAR	404	80
476	GGAGAGGATG	87	TTSNLRR	193	RSDNLAR	299	QSDHLTR	405	80
477	GGAGAGGATG	88	TTSNLRR	194	RSDNLAR	300	QRAHLAR	406	100
478	GGAGAGGATG	89	TTSNLRR	195	RSDNLAR	301	QSGHLRR	407	60
479	GTGGCGGACC	90	DSSNLTR	196	RSDELQR	302	RSDALAR	408	8.5
480	GTGGCGGACC	91	DSSNLTR	197	RADTLRR	303	RSDALAR	409	5
483	GAGGGCGAAG	92	QSANLAR	198	ESSKLKR	304	RSDNLAR	410	130
484	GAGGGCGAAG	93	QSDNLAR	199	ESSKLKR	305	RSDNLAR	411	1000
485	GGAGAGGTTT	94	QSSALAR	200	RSDNLAR	306	QRAHLAR	412	110
487	GGAGAGGTTT	95	NRATLAR	201	RSDNLAR	307	QSGHLAR	413	76.9
488	TGGTAGGGGG	96	RSDHLAR	202	RSDNLTT	308	RSDHLTT	414	35
490	TAGGGGGTGG	97	RSDSLLR	203	RSDHLTR	309	RSDNLTT	415	1.5
503	GCCGAGGTGC	98	RSDSLLR	204	RSDNLAR	310	ERGTLAR	416	50
504	GCCGAGGTGC	99	RSDSLLR	205	RSDNLAR	311	DRSDLTR	417	25
505	GCCGAGGTGC	100	RSDSLLR	206	RSDNLAR	312	DCRDLAR	418	65
526	GCGGGCGGGC	101	RSDHLTR	207	ERGHLTR	313	RSDTLKK	419	8
543	GAGTGTGTGA	102	RSDLLQR	208	MSHHLKE	314	RSDHLSR	420	50
544	GAGTGTGTGA	103	RSDSLLR	209	MSHHLKE	315	RSDNLAR	421	125
545	GAGTGTGTGA	104	RKDSLVR	210	TSDHLAS	316	RSDNLTR	422	32
546	GAGTGTGTGA	105	RSDLLQR	211	MSHHLKT	317	RLDGLRT	423	500
547	GAGTGTGTGA	106	RKDSLVR	212	TSGHLTS	318	RSDNLTR	424	500
548	GAGTGTGTGA	107	RSSLLQR	213	MSHHLKT	319	RSDHLSR	425	500
549	GAGTGTGTGA	108	RSSLLQR	214	MSHHLKE	320	RSDHLSR	426	500
550	GAGTGTGTGA	109	RKDSLVR	215	TKDHLAS	321	RSDNLTR	427	20

551	GAGTGTGTGA	110	RSDLLQR	216	MSHHLKT	322	RSDHLSR	428	50
552	GAGTGTGTGA	111	RKDSLVR	217	MSHHLKT	323	RSDNLTR	429	31
553	GAGTGTGTGA	112	RSDSLLR	218	MSHHLKE	324	RSDNLTR	430	125
554	GAGTGTGTGA	113	RKDSLVR	219	TSDHLAS	325	RSDNLAR	431	62.5
558	TGCGGGGCA	114	QSGDLTR	220	RSDHLTR	326	DSGHLAS	432	21
559	GAGTGTGTGA	115	RSDSLLR	221	TSDHLAS	327	RSDNLAR	433	1000
560	GAGTGTGTGA	116	RSSLLQR	222	MSHHLKT	328	RSDHLSR	434	500
561	GAGTGTGTGA	117	RKDSLVR	223	MSHHLKE	329	RSDNLAR	435	1000
562	GAGTGTGTGA	118	RSDSLLR	224	TSGHLTS	330	RSDNLAR	436	1000
565	GATGCTGAG	119	RSDNLTR	225	TSSELQR	331	QQSNLAR	437	100
567	GAAGATGAC	120	EKANLTR	226	TSANLSR	332	QRSNLVR	438	47.5
568	GATGACGAC	121	EKANLTR	227	DSSNLTR	333	TSANLSR	439	300
569	GTAGTTGTG	122	RSDSLLR	228	TGGSLAR	334	QRSALTR	440	52

TABLE 2

<u>SBS#</u>		<u>SEQ</u>		<u>SEQ</u>		<u>SEQ</u>		<u>SEQ</u>	<u>Kd</u>
	TARGET	ID	F1	ID	F2	ID	F3	ID	(nM)
201	GCAGCCTTG	441	RSDSLTS	646	ERSTLTR	851	QRADLRR	1056	1000
202	GCAGCCTTG	442	RSDSLTS	647	ERSTLTR	852	QRADLAR	1057	1000
203	GCAGCCTTG	443	RSDSLTS	648	ERSTLTR	853	QRATLRR	1058	1000
204	GCAGCCTTG	444	RSDSLTS	649	ERSTLTR	854	QRATLAR	1059	1000
205	GAGGTAGAA	445	QSANLAR	650	QSATLAR	855	RSDNLSR	1060	80
206	GAGGTAGAA	446	QSANLAR	651	QSAVLAR	856	RSDNLSR	1061	1000
207	GAGTGGTTA	447	QRASLAS	652	RSDHLTT	857	RSDNLAR	1062	70
208	TAGGTCTTA	448	QRASLAS	653	DRSALAR	858	RSDNLAS	1063	1000
209	GGAGTGGTT	449	QSSALAR	654	RSDALAR	859	QRAHLAR	1064	35
210	GGAGTGGTT	450	NRDTLAR	655	RSDALAR	860	QRAHLAR	1065	65
211	GGAGTGGTT	451	QSSALAR	656	RSDALAS	861	QRAHLAR	1066	140
212	GGAGTGGTT	452	NRDTLAR	657	RSDALAS	862	QRAHLAR	1067	400
213	GTTGCTGGA	453	QRAHLAR	658	QSSTLAR	863	QSSALAR	1068	1000
214	GTTGCTGGA	454	QRAHLAR	659	QSSTLAR	864	NRDTLAR	1069	1000
215	GAAGTCTGT	455	NRDHLMV	660	DRSALAR	865	QSANLSR	1070	1000
216	GAAGTCTGT	456	NRDHLTT	661	DRSALAR	866	QSANLSR	1071	1000
217	GAGGTCGTA	457	QRSALAR	662	DRSALAR	867	RSDNLAR	1072	40
219	GATGTTGAT	458	QQSNLAR	663	NRDTLAR	868	NRDNLSR	1073	1000
220	GATGTTGAT	459	QQSNLAR	664	NRDTLAR	869	QQSNLSR	1074	1000
221	GATGAGTAC	460	DRSNLRT	665	RSDNLAR	870	NRDNLAR	1075	1000
222	GATGAGTAC	461	ERSNLRT	666	RSDNLAR	871	NRDNLAR	1076	1000
223	GATGAGTAC	462	DRSNLRT	667	RSDNLAR	872	QQSNLAR	1077	105
224	GATGAGTAC	463	ERSNLRT	668	RSDNLAR	873	QQSNLAR	1078	1000
225	TGGGAGGTC	464	DRSALAR	669	RSDNLAR	874	RSDHLTT	1079	6
226	GCAGCCTTG	465	RGDALTS	670	ERGTTLAR	875	QSGSLTR	1080	1000
227	GCAGCCTTG	466	RGDALTV	671	ERGTTLAR	876	QSGSLTR	1081	1000
228	GCAGCCTTG	467	RGDALTM	672	ERGTTLAR	877	QSGSLTR	1082	1000
229	GCAGCCTTG	468	RGDALTS	673	ERGTTLAR	878	RSDELTR	1083	1000

230	GCAGCCTTG	469	RGDALTV	674	ERGTLLAR	879	RSDELTR	1084	1000
231	GCAGCCTTG	470	RGDALTM	675	ERGTLLAR	880	RSDELTR	1085	1000
232	GGTGTGGTG	471	RSDALTR	676	RSDALAR	881	NRSHLAR	1086	50
233	GGTGTGGTG	472	RSDALTR	677	RSDALAR	882	QASHLAR	1087	100
235	GTAGAGGTG	473	RSDALTR	678	RSDNLAR	883	QRGALAR	1088	80
236	GGGGAGGGG	474	RSDHLAR	679	RSDNLAR	884	RSDHLSR	1089	0.3
237	GGGGAGGCC	475	ERGTLLAR	680	RSDNLAR	885	RSDHLSR	1090	0.3
238	GGGGAGGCC	476	ERGTLLAR	681	RSDNLQR	886	RSDHLSR	1091	0.8
239	GGCGGGGAG	477	RSDNLTR	682	RSDHLTR	887	DRSHLAR	1092	0.4
240	GCAGGGGAG	478	RSDNLTR	683	RSDHLSR	888	QSGSLTR	1093	1
242	GGGGGTGCT	479	QSSDLRR	684	QSSHLAR	889	RSDHLSR	1094	1
243	GTGGGCGCT	480	QSSDLRR	685	DRSHLAR	890	RSDALAR	1095	75
244	TAAGAAGGG	481	RSDHLAR	686	QSGNLTR	891	QSGNLRT	1096	100
245	TAAGAAGGG	482	RSDHLAR	687	QSANLTR	892	QSGNLRT	1097	235
246	GAAGGGGAG	483	RSDNLAR	688	RSDHLAR	893	QSGNLTR	1098	2
247	GAAGGGGAG	484	RSDNLAR	689	RSDHLAR	894	QSGNLRR	1099	2
276	GCGGCCGCG	485	RSDELTR	690	ERGTLLAR	895	RSDEKRR	1100	90
277	GCGGCCGCG	486	RSDELTR	691	DRSSLTR	896	RSDEKRR	1101	107
278	GCGGCCGCG	487	QSWELTR	692	ERGTLLAR	897	RSDEKRR	1102	190
279	GCGGCCGCG	488	QSWELTR	693	DRSSLTR	898	RSDEKRR	1103	260
280	GCGGCCGCG	489	QSGSLTR	694	ERGTLLAR	899	RSDEKRR	1104	160
281	GCGGCCGCG	490	QSGSLTR	695	DRSSLTR	900	RSDEKRR	1105	225
282	GCAGAAGTG	491	RGDALTR	696	QSANLTR	901	QSADLAR	1106	1000
283	GCAGAAGTG	492	RSDALTR	697	QSGNLTR	902	QSGSLTR	1107	2
284	GCGGCCGCG	493	QSGSLTR	698	RSDHLTT	903	RSDEKRR	1108	1000
285	TGTGCGGCC	494	ERGTLLAR	699	RSDELTR	904	SRDHLQS	1109	1000
287	GCAGAAGCG	495	RGPDLAR	700	QSANLTR	905	QSGSLTR	1110	1000
288	GCAGAAGCG	496	RGPDLAR	701	QSANLTR	906	QSGSLTR	1111	1000
289	GCAGAAGCG	497	RGPDLAR	702	QSGNLQR	907	QSGSLTR	1112	800
290	GCAGAAGCG	498	RSDELAR	703	QSANLQR	908	QSADLAR	1113	1000
292	GCAGAAGCG	499	RSDELTR	704	QSANLQR	909	QSGSLTR	1114	1000
293	GTGTGCGGC	500	DRSHLTR	705	ERHSLQT	910	RSDALTR	1115	320
296	TGCGCGGCC	501	ERGTLLAR	706	RSDELTR	911	DRDHLQS	1116	1000

297	TGCGCGGCC	502	ERGTLLAR	707	RSDELRR	912	DRSHLQT	1117	500
298	GCTTAGGCA	503	QTGELRR	708	RSDNLQK	913	TSGDLSR	1118	4000
299	GCTTAGGCA	504	QTSDLRR	709	RSDNLQK	914	QSSDLQR	1119	4000
300	GCTTAGGCA	505	QTADLRR	710	RSDNLQR	915	QSSDLSR	1120	400
301	GCTTAGGCA	506	QSADLRR	711	RSDNLQT	916	QSSDLSR	1121	350
302	GCTTAGGCA	507	QSGSLTR	712	RSDNLQT	917	QSSDLSR	1122	75
303	GCTTAGGCA	508	QTGSLTR	713	RSDNLQT	918	QSSDLSR	1123	135
304	GCTTAGGCA	509	QTADLTR	714	RSDNLQT	919	QSSDLSR	1124	230
305	GCTTAGGCA	510	QTGDLTR	715	RSDNLQT	920	QSSDLSR	1125	230
306	GCTTAGGCA	511	QTASLTR	716	RSDNLQT	921	QSSDLSR	1126	280
307	GAAGAAGCG	512	RSDELRR	717	QSGNLQR	922	QSGNLSR	1127	50.5
308	GAAGAAGCG	513	RSDELRR	718	QSANLQR	923	QSANLQR	1128	1000
309	GGAGATGCC	514	ERSDLRR	719	QSSNLQR	924	QSGHLSR	1129	4000
310	GGAGATGCC	515	DRSDLTR	720	NRDNLQT	925	QSGHLSR	1130	1000
311	GGAGATGCC	516	DRSTLTR	721	NRDNLQR	926	QSGHLSR	1131	170
312	GGAGATGCC	517	ERGTLLAR	722	NRDNLQR	927	QSGHLSR	1132	2000
313	GGAGATGCC	518	DRSDLTR	723	QRSNLQR	928	QSGHLSR	1133	1000
314	GGAGATGCC	519	DRSSLTR	724	QSSNLQR	929	QSGHLSR	1134	117.5
315	GGAGATGCC	520	ERGTLLAR	725	QSSNLQR	930	QSGHLSR	1135	265
316	GGAGATGCC	521	ERGTLLAR	726	QRDNLQR	931	QSGHLSR	1136	3000
318	TAGGAGATGC	522	RSDALTS	727	RSDNLAR	932	RSDNLAS	1137	100
319	GGGAAGGG	523	KTSHLRA	728	QSGNLSR	933	RSDHLSR	1138	125
320	GGGAAGGG	524	RSDHLTR	729	QSGNLSR	934	RSDHLSR	1139	5
321	GGCGGAGAT	525	TTSNLRR	730	QSGHLQR	935	DRSHLTR	1140	200
323	GGCGGAGAT	526	TTSNLRR	731	QSGHLQR	936	DRDHLTR	1141	600
324	GGCGGAGAT	527	TTSNLRR	732	QSGHLQR	937	DRDHLTR	1142	200
325	GTATCTGCT	528	NSSDLTR	733	NSDVLTS	938	QSDVLTR	1143	1000
326	GTATCTGTT	529	NSDALTR	734	NSDVLTS	939	QSDVLTR	1144	1000
327	TCTGCTGGG	530	RSDHLTR	735	NSADLTR	940	NSDDLTR	1145	1000
328	TCTGTTGGG	531	RSDHLTR	736	NSSALTS	941	NSDDLTR	1146	1000
349	GGTGTGCC	532	DCRDLAR	737	DSGSLTR	942	TSGHLTR	1147	1000
350	TCCGAGGGT	533	TSGHLTR	738	RSDNLTR	943	DCRDLTT	1148	332
351	GCTGGTGTC	534	DSGSLTR	739	TSGHLTR	944	TLHTLTR	1149	1000

352	GGAGGGGTG	535	RSDSLLR 740	RSDHLTR 945	QSDHLTR 1150	26
353	GTTGGAGCC	536	DCRDLAR 741	QSDHLTR 946	TSGALTR 1151	1000
354	GAAGAGGAC	537	DSSNLTR 742	RSDNLTR 947	QRSNLVR 1152	28
355	GAAGAGGAC	538	EKANLTR 743	RSDNLTR 948	QRSNLVR 1153	20
356	GGCTGGGCG	539	RSDELRR 744	RSDHLTK 949	DSDHLSR 1154	1000
357	GGCTGGGCG	540	RSDELRR 745	RSDHLTK 950	DSDHLSR 1155	1000
358	GGCTGGGCG	541	RSDELRR 746	RSDHLTK 951	DSSHLSR 1156	225
361	GGGTTTGGG	542	RSDHLTR 747	QSSALTR 952	RSDHLTR 1157	130
363	GGGTTTGGG	543	RSDHLTR 748	QSSVLTR 953	RSDHLTR 1158	200
364	GTGTCCGAAG	544	RSDNLTR 749	DSAVLTT 954	RSDSLTR 1159	1000
365	GGTGCTGGT	545	QASHLTR 750	QASVLTR 955	QASHLTR 1160	600
366	GAGGGTGCT	546	QASVLTR 751	QASHLTR 956	RSDNLTR 1161	1000
367	GGGGGCGGG	547	RSDHLTR 752	DSGHLTR 957	RSDHLQR 1162	60
368	GAGGGGGCG	548	RSDELTR 753	RSDHLTR 958	RSDNLTR 1163	3.5
369	GTAGTTGTG	549	RSDALTR 754	TGGSLAR 959	QSGSLTR 1164	95
370	GTAGTTGTG	550	RSDALTR 755	NRATLAR 960	QSASLTR 1165	300
371	GTAGTTGTG	551	RSDALTR 756	NRATLAR 961	QSGSLTR 1166	175
372	GTAGTTGTG	552	RSDSLLR 757	TGGSLAR 962	QSASLTR 1167	112.5
373	GTAGTTGTG	553	RSDSLLR 758	NRATLAR 963	QSASLTR 1168	320
374	GCTGAGGAA	554	QRSNLVR 759	RSDNLTR 964	TSSELQR 1169	3.3
375	GAGGAAGAT	555	QQSNLAR 760	QSGNLQR 965	RSDNLTR 1170	85
377	GTGTTGGCAG	556	QSGSLTR 761	RGDALTS 966	RSDALTR 1171	89
378	GCCGAGGAGA	557	RSDNLTR 762	RSDNLTR 967	DRSSLTR 1172	31
379	GCCGAGGAGA	558	RSDNLTR 763	RSDNLTR 968	ERGTLAR 1173	3
380	GAGTCGGAAG	559	QSANLAR 764	RSDELTT 969	RSDNLAR 1174	1000
381	GCAGCTGCGC	560	RSDELTR 765	QSSDLQR 970	QSGDLTR 1175	1.5
383	TGTTTGGTAT	561	QSATLAR 766	RGDALTS 971	RSDHLTT 1176	1000
384	GTGGGCTTCA	562	DRSALTT 767	DRSHLAR 972	RSDALAR 1177	60
385	GGGGCGGAGC	563	RSDNLTR 768	RSDTLKK 973	RSDHLSR 1178	1.2
386	GGGGCGGAGC	564	RSDNLTR 769	RSDELQR 974	RSDHLSR 1179	0.4
387	GGCGAGGCAA	565	QSGSLTR 770	RSDNLAR 975	DRSHLAR 1180	2.5
388	GGCGAGGCAA	566	QSGDLTR 771	RSDNLAR 976	DRSHLAR 1181	28
390	GTGGCAGCGG	567	RSDTLKK 772	QSSDLQK 977	RSDALAR 1182	20

392	GTGGCAGCGG	568	RSDLTR	773	QSSDLQK	978	RSDALAR	1183	1000
396	GCGGGAGCAG	569	QSGSLTR	774	QSGHLQR	979	RSDTLKK	1184	18.8
397	GCGGGAGCAG	570	QSGDLTR	775	QSGHLQR	980	RSDTLKK	1185	25
400	TCAGTGGTGG	571	RSDALAR	776	RSDSLAR	981	QSGDLRT	1186	40
405	GCGGCCGCA	572	RSDLTR	777	ERGTAR	982	RSDERKR	1187	110
406	GCGGCCGCA	573	RSDLTR	778	DRSSLTR	983	RSDERKR	1188	110
407	GCGGCCGCA	574	QSWELTR	779	ERGTAR	984	RSDERKR	1189	410
408	GCGGCCGCA	575	QSWELTR	780	DRSSLTR	985	RSDERKR	1190	380
409	GCGGCCGCA	576	QSGSLTR	781	ERGTAR	986	RSDERKR	1191	50
410	GCAGAAGTC	577	RSDLTR	782	QSGNLTR	987	QSGSLTR	1192	3
411	GCGGCCGCA	578	QSGSLTR	783	RSDHLTT	988	RSDERKR	1193	1000
412	GCGTGGGCG	579	QSGSLTR	784	RSDHLTT	989	RSDERKR	1194	5
413	GCGTGGGCA	580	QSGSLTR	785	RSDHLTT	990	RSDERKR	1195	5
414	GCAGAAGCA	581	RSDLTR	786	QSANLQR	991	QSGSLTR	1196	1000
415	GTGTGCGGA	582	DRSHLTR	787	ERHSLQT	992	RSDLTR	1197	1000
416	TGTGCGGCC	583	ERGTAR	788	RSDELRR	993	DRSHLQT	1198	1000
493	GGGGTGGCGG	584	RSDTLKK	789	RSDSLAR	994	RSDHLR	1199	300
494	GCCGAGGAGA	585	RSDNLTR	790	RSDNLTR	995	DRSSLTR	1200	90
496	GGTGGTGGC	586	DTSHLRR	791	TSGHLQR	996	TSGHLR	1201	1000
497	GTTTGCGTC	587	ETASLRR	792	DSHLQR	997	TSSALR	1202	1000
498	GAAGAGGCA	588	QTGELRR	793	RSDNLQR	998	QSGNLR	1203	30
499	GCTTGTGAG	589	RTSNLRR	794	TSSHLQK	999	DTDHLRR	1204	1000
500	GCTTGTGAG	590	RSDNLTR	795	QSSNLQT	1000	DRSHLAR	1205	1000
501	GTGGGGGTT	591	NRATLAR	796	RSDHLR	1001	RSDALAR	1206	8
502	GGGGTGGGA	592	QSAHLAR	797	RSDALAR	1002	RSDHLR	1207	60
507	GAGGTAGAGG	593	RSDNLAR	798	QRSALAR	1003	RSDNLAR	1208	10
508	GAGGTAGAGG	594	RSDNLAR	799	QSATLAR	1004	RSDNLAR	1209	10
509	GTCGTGTGGC	595	RSDHLTT	800	RSDALAR	1005	DRSALAR	1210	100
510	GTTGAGGAAG	596	QSGNLAR	801	RSDNLAR	1006	NRATLAR	1211	100
511	GTTGAGGAAG	597	QSGNLAR	802	RSDNLAR	1007	QSSALAR	1212	100
512	GAGGTGGAAG	598	QSGNLAR	803	RSDALAR	1008	RSDNLAR	1213	10
513	GAGGTGGAAG	599	QSANLAR	804	RSDALAR	1009	RSDNLAR	1214	1.5
514	TAGGTGGTGG	600	RSDLTR	805	RSDALAR	1010	RSDNLTT	1215	10

515	TGGGAGGAGT	601	RSDNLTR	806	RSDNLTR	1011	RSDHLTT	1216	0.5
516	GGAGGAGCT	602	TTSELRR	807	QSGHLQR	1012	QSGHLSR	1217	700
517	GGAGCTGGGG	603	RTDHLRR	808	TSSELQR	1013	QSGHLSR	1218	50
518	GGGGGAGGAG	604	QTGHLRR	809	QSGHLQR	1014	RSDHLSR	1219	30
519	GGGGAGGAGA	605	RSDNLAR	810	RSDNLSR	1015	RSDHLSR	1220	0.3
520	GGAGGAGAT	606	TTANLRR	811	QSGHLQR	1016	QSGHLSR	1221	300
521	GCAGCAGGA	607	QTGHLRR	812	QSGELQR	1017	QSGELSR	1222	1000
522	GATGAGGCA	608	QTGELRR	813	RSDNLQR	1018	TSANLSR	1223	200
527	GGGGAGGATC	609	TTSNLRR	814	RSSNLQR	1019	RSDHLSR	1224	2
528	GGGGAGGATC	610	TTSNLRR	815	RSSNLQR	1020	RSDHLSR	1225	10
529	GAGGCTTGGG	611	RTDHLRK	816	TSAELQR	1021	RSSNLSR	1226	1000
531	GCGGAGGCTT	612	TTGELRR	817	RSSNLQR	1022	RSELSR	1227	160
532	GCGGAGGCTT	613	QSSDLQR	818	RSSNLQR	1023	RSELSR	1228	100
533	GCGGAGGCTT	614	QSSDLQR	819	RSDNLAR	1024	RSADLSR	1229	7
534	GCGGAGGCTT	615	QSSDLQR	820	RSDNLAR	1025	RSDDLRR	1230	10
535	GCAGCCGGG	616	RTDHLRR	821	ESSDLQR	1026	QSGELSR	1231	1000
538	GCAGAGGCTT	617	QSSDLQR	822	RSDNLAR	1027	QSGSLTR	1232	70
540	TGGGCAGGCC	618	DRSHLTR	823	QSGSLTR	1028	RSDHLTT	1233	55
541	GGGGAGGAT	619	TTSNLRR	824	RSSNLQR	1029	RSDHLSR	1234	3
570	GGGGAAGGCT	620	DSGHLTR	825	QRSNLVR	1030	RSDHLTR	1235	20
571	GTGTGTGTGT	621	RSDSLTR	826	QRSNLVR	1031	RSDSLLR	1236	1000
572	GCATACGTGG	622	RSDSLLR	827	DKGNLQS	1032	QSDDLTR	1237	1000
573	GCATACGTG	623	RSDSLLR	828	DKGNLQS	1033	QSGDLTR	1238	1000
574	TACGTGGGGT	624	RSDHLTR	829	RSDHLTR	1034	DKGNLQT	1239	25
575	TACGTGGGCT	625	DFSHLTR	830	RSDHLTR	1035	DKGNLQT	1240	472
576	GAGGGTGTG	626	NSDTLAR	831	TSGHLTR	1036	RSDNLTR	1241	200
577	GGAGCGGGGA	627	RSDHLSR	832	RSDELQR	1037	QSDHLTR	1242	200
579	GGGGTTGAGG	628	RSDNLTR	833	NRDTLAR	1038	TSGHLTR	1243	200
580	GGTGTGAG	629	QRAHLAR	834	NRDTLAR	1039	TSGHLTR	1244	1000
581	TACGTGGGTT	630	QSSHLTR	835	RSDSLLR	1040	DKGNLQT	1245	382
583	GTAGGGGTTG	631	NSSALTR	836	RSDHLTR	1041	QSASLTR	1246	46
584	GAAGGCGGAG	632	QAGHLTR	837	DKSHLTR	1042	QSGNLTR	1247	1000
585	GAAGGCGGAG	633	QAGHLTR	838	DSGHLTR	1043	QSGNLTR	1248	1000

587	GGGGGTTACG	634	DKGNLQT	839	TSGHLTR	1044	RSDHLSK	1249	500
588	GGGGGGGGGG	635	RSDHLSR	840	RSDHLTR	1045	RSDHLSK	1250	30
589	GGAGTATGCT	636	DSGHLAS	841	QSATLAR	1046	QSDHLTR	1251	1000
595	TGGTTGGTAT	637	QRGSLAR	842	RGDALTR	1047	RSDHLTT	1252	73.3
597	TGGTTGGTA	638	QNSAMRK	843	RGDALTS	1048	RSDHLTT	1253	1000
598	TGGTTGGTA	639	QRGSLAR	844	RDGSLTS	1049	RSDHLTT	1254	1000
599	TGGTTGGTA	640	QNSAMRK	845	RDGSLTS	1050	RSDHLTT	1255	1000
600	GAGTCGGAA	641	QSANLAR	846	RSDELRT	1051	RSDNLAR	1256	206.7
601	GAGTCGGAA	642	RSANLTR	847	RLDGLRT	1052	RSDNLAR	1257	606.7
602	GAGTCGGAA	643	RSANLTR	848	RQDTLVG	1053	RSDNLAR	1258	616.7
603	GAGTCGGAA	644	QSGNLAR	849	RSDELRT	1054	RSDNLAR	1259	166.7
606	GGGGAGGATC	645	TTSNLRR	850	RSDNLQR	1055	RSDHLSR	1260	0.2

TABLE 3

<u>SBS#</u>	<u>TARGET</u>	<u>SEQ</u> <u>ID</u>	<u>F1</u>	<u>SEQ</u> <u>ID</u>	<u>F2</u>	<u>SEQ</u> <u>ID</u>	<u>F3</u>	<u>SEQ</u> <u>ID</u>	<u>Kd</u> <u>(nM)</u>
897	GAGGAGGTGA	1261	RSDALAR	1347	RSDNLAR	1433	RSDNLVR	1519	0.07
828	GCGGAGGACC	1262	EKANLTR	1348	RSDNLAR	1434	RSDERKR	1520	0.1
884	GAGGAGGTGA	1263	RSDSLTR	1349	RSDNLAR	1435	RSDNLVR	1521	0.15
817	GAGGAGGTGA	1264	RSDSLTR	1350	RSDNLAR	1436	RSDNLAR	1522	0.31
666	GCGGAGGCGC	1265	RSDDLTR	1351	RSDNLTR	1437	RSDTLKK	1523	0.5
829	GCGGAGGACC	1266	EKANLTR	1352	RSDNLAR	1438	RSDTLKK	1524	0.52
670	GACGTGGAGG	1267	RSDNLAR	1353	RSDALAR	1439	DRSNLTR	1525	0.57
801	AAGGAGTCGC	1268	RSADLRT	1354	RSDNLAR	1440	RSDNLTQ	1526	0.85
668	GTGGAGGCCA	1269	ERGTLAR	1355	RSDNLAR	1441	RSDALAR	1527	1.13
895	ATGGATTTCAG	1270	QSHDLTK	1356	TSGNLVR	1442	RSDALTQ	1528	1.4
799	GGGGGAGCTG	1271	QSSDLQR	1357	QRAHLER	1443	RSDHLSR	1529	1.85
798	GGGGGAGCTG	1272	QSSDLQR	1358	QSGHLQR	1444	RSDHLSR	1530	3
842	GAGGTGGGCT	1273	DRSHLTR	1359	RSDALAR	1445	RSDNLAR	1531	5.4
894	TCAGTGGTAT	1274	QRSALAR	1360	RSDALSR	1446	QSHDLTK	1532	6.15
892	ATGGATTTCAG	1275	QSHDLTK	1361	QSNLVR	1447	RSDALTQ	1533	6.2
888	TCAGTGGTAT	1276	QSSSLVR	1362	RSDALSR	1448	QSHDLTK	1534	14
739	GCGGGCGGGC	1277	RSDHLTR	1363	ERHGLTR	1449	RSDDLRR	1535	16.5
850	CAGGCTGTGG	1278	RSDALTR	1364	QSSDLTR	1450	RSDNLRE	1536	17
797	GCAGAGGCTG	1279	QSSDLQR	1365	RSDNLAR	1451	QSGDLTR	1537	17.5
891	TCAGTGGTAT	1280	QSSSLVR	1366	RSDALSR	1452	QSGSLRT	1538	18.5
887	TCAGTGGTAT	1281	QRSALAR	1367	RSDALSR	1453	QSGDLRT	1539	23.75
672	TCGGACGTGG	1282	RSDALAR	1368	DRSNLTR	1454	RSDELRT	1540	24
836	GGGGAGGCC	1283	ERGTLAR	1369	RSDNLAR	1455	RSDHLSR	1541	24.25
674	GCGGCGTCGG	1284	RSDELRT	1370	RADTLRR	1456	RSDTLKK	1542	27.5
849	GGGGCCCTGG	1285	RSDALRE	1371	DRSSLTR	1457	RSDHLTQ	1543	29.05
825	GAATGGGCAG	1286	QSGSLTR	1372	RSDHLTT	1458	QSGNLTR	1544	37.3
673	GCGGGTGTCT	1287	DRSALAR	1373	QSSHLAR	1459	RSDTLKK	1545	48.33
848	GGGGAGGCC	1288	DRSSLTR	1374	RSDNLAR	1460	RSDHLSR	1546	49.5

662	AGAGCGGCAC	1289	QTGSLTR	1375	RSDELQR	1461	QSGHLNQ	1547	50
667	GAGTCGGACG	1290	DRSNLTR	1376	RSDELRT	1462	RSDNLAR	1548	50
803	GCAGCGGCTC	1291	QSSDLQR	1377	RSDELQR	1463	QSGSLTR	1549	57.5
671	TCGGACGAGT	1292	RSDNLAR	1378	DRSNLTR	1464	RSDELRT	1550	64
851	GAGATGGATC	1293	QSSNLQR	1379	RRDVL MN	1465	RLHNLQR	1551	74
804	GCAGCGGCTC	1294	QSSDLQR	1380	RSDDLNR	1466	QSGSLTR	1552	82.5
669	GACGAGTCGG	1295	RSDELRT	1381	RSDNLAR	1467	DRSNLTR	1553	90
682	GCTGCAGGAG	1296	RSDHLAR	1382	QSGDLTR	1468	QSSDLSR	1554	90
845	GAGATGGATC	1297	QSSNLQR	1383	RSDALRQ	1469	RLHNLQR	1555	112.5
663	AGAGCGGCAC	1298	QTGSLTR	1384	RSDELQR	1470	KNWKLQA	1556	115
738	GCGGGGTCCG	1299	ERGTLTT	1385	RSDHLSR	1471	RSDDLRR	1557	120
664	AGAGCGGCAC	1300	QTGSLTR	1386	RADTLRR	1472	ASSRLAT	1558	125
833	GACTAGGACC	1301	EKANLTR	1387	RSDNLTK	1473	DRSNLTR	1559	136
685	GCTGCAGGAG	1302	RSDHLAR	1388	QSGSLTR	1474	QSSDLSR	1560	150
835	TAGGGAGCGT	1303	RADTLRR	1389	QSGHLTR	1475	RSDNLTT	1561	150
847	TAGGGAGCGT	1304	RSDDLTR	1390	QSGHLTR	1476	RSDNLTT	1562	150
818	GAATGGGCAG	1305	QSGSLTR	1391	RSDHLTT	1477	QSSNLVR	1563	167
834	GACTAGGACC	1306	EKANLTR	1392	RSDHLTT	1478	DRSNLTR	1564	186
837	GGGGCCCTGG	1307	RSDALRE	1393	DRSSLTR	1479	RSDHLSR	1565	222
764	GCAGAGGCTG	1308	TSGELVR	1394	RSDNLAR	1480	QSGDLTR	1566	255
774	GCAGCGGTAG	1309	QRSALAR	1395	RSDELQR	1481	QSGDLTR	1567	258
765	GCCGAGGCCG	1310	ERGTLAR	1396	RSDNLAR	1482	ERGTLAR	1568	262.5
766	GCCGAGGCCG	1311	ERGTLAR	1397	RSDNLAR	1483	DRSDLTR	1569	262.5
775	GCAGCGGTAG	1312	QSGALTR	1398	RSDELQR	1484	QSGDLTR	1570	265
763	GCAGAGGCTG	1313	TSGELVR	1399	RSDNLAR	1485	QSGSLTR	1571	275
838	GGGGCCCTGG	1314	RSDALRE	1400	DRSSLTR	1486	RSDHLTA	1572	300
841	GAGTGTGAGG	1315	RSDNLAR	1401	QSSHLAS	1487	RSDNLAR	1573	300
770	TTGGCAGCCT	1316	DRSSLTR	1402	QSGSLTR	1488	RSDSLTK	1574	325
767	GGGGGAGCTG	1317	QSSDLAR	1403	QSGHLQR	1489	RSDHLSR	1575	335
800	TTGGCAGCCT	1318	ERGTLAR	1404	QSGSLTR	1490	RSDSLTK	1576	400
832	GACTAGGACC	1319	EKANLTR	1405	RSDNLTT	1491	DRSNLTR	1577	408
844	GAGATGGATC	1320	QSSNLQR	1406	RSDALRQ	1492	RSDNLQR	1578	444
683	GCTGCAGGAG	1321	QSGHLAR	1407	QSGSLTR	1493	QSSDLSR	1579	500

805	GCAGCGGTAG 1322	QRSALAR 1408	RSDELQR 1494	QSGSLTR 1580	500
839	GAGTGTGAGG 1323	RSDNLAR 1409	TSDHLAS 1495	RSDNLAR 1581	625
840	GAGTGTGAGG 1324	RSDNLAR 1410	MSHHLKT 1496	RSDNLAR 1582	625
830	GGAGAGTCGG 1325	RSDELRT 1411	RSDNLAR 1497	QRAHLAR 1583	683
831	GGAGAGTCGG 1326	RSDDLTK 1412	RSDNLAR 1498	QRAHLAR 1584	700
684	GCTGCAGGAG 1327	RSAHLAR 1413	QSGSLTR 1499	QSSDLSR 1585	850
846	GAGATGGATC 1328	QSSNLQR 1414	RRDVLMN 1500	RSDNLQR 1586	889.5
819	AAGTAGGGTG 1329	QSSHLTR 1415	RSDNLTT 1501	RSDNLTQ 1587	1000
820	ACGGTAGTTA 1330	QSSALTR 1416	QRSALAR 1502	RSDTLTQ 1588	1000
821	ACGGTAGTTA 1331	NRATLAR 1417	QRSALAR 1503	RSDTLTQ 1589	1000
822	GTGTGCTGGT 1332	RSDHLTT 1418	ERQHLAT 1504	RSDALAR 1590	1000
823	GTGTGCTGGT 1333	RSDHLTK 1419	ERQHLAT 1505	RSDALAR 1591	1000
824	GTGTGCTGGT 1334	RSDHLTT 1420	DRSHLRT 1506	RSDALAR 1592	1000
885	GTGTGCTGGT 1335	RSDHLTK 1421	DRSHLRT 1507	RSDALAR 1593	1000
886	TCAGTGGTAT 1336	QSSSLVR 1422	RSDALSR 1508	QSGDLRT 1594	1000
889	ATGGATTCAG 1337	QSGSLTT 1423	QSSNLVR 1509	RSDALTQ 1595	1000
890	CTGGTATGTC 1338	QRSHLTT 1424	QRSALAR 1510	RSDALRE 1596	1000
896	AAGTAGGGTG 1339	TSGHLVR 1425	RSDNLTT 1511	RSDNLTQ 1597	1000
898	ACGGTAGTTA 1340	NRATLAR 1426	QSSSLVR 1512	RSDTLTQ 1598	1000
899	CTGGTATGTC 1341	QRSHLTT 1427	QSSSLVR 1513	RSDALRE 1599	1000
900	CTGGTATGTC 1342	MSHHLKE 1428	QSSSLVR 1514	RSDALRE 1600	1000
901	CTGGTATGTC 1343	MSHHLKE 1429	QRSALAR 1515	RSDALRE 1601	1000
773	GCAGCGGTAG 1344	QSGALTR 1430	RSDELQR 1516	QSGSLTR 1602	1250
768	GGGGGAGCTG 1345	QSSDLAR 1431	QRAHLER 1517	RSDHLR 1603	2000
681	GCTGCAGGAG 1346	RSAHLAR 1432	QSGDLTR 1518	QSSDLSR 1604	3000

TABLE 4

<u>SBS#</u>	<u>TARGET</u>	<u>SEQ</u> <u>ID</u>	<u>F1</u>	<u>SEQ</u> <u>ID</u>	<u>F2</u>	<u>SEQ</u> <u>ID</u>	<u>F3</u>	<u>SEQ</u> <u>ID</u>	<u>Kd</u> <u>(nM)</u>
607	AAGGTGGCAG	1605	QSGDLTR	1707	RSDSLAR	1809	RLDNRTA	1911	6.5
608	TTGGCTGGGC	1606	GSWHLTR	1708	QSSDLQR	1810	RSDSLTK	1912	8
611	GTGGCTGCAG	1607	QSGDLTR	1709	QSSDLQR	1811	RSDALAR	1913	11.5
612	GTGGCTGCAG	1608	QSGTLTR	1710	QSSDLQR	1812	RSDALAR	1914	0.38
613	TTGGCTGGGC	1609	RSDHLAR	1711	QSSDLQR	1813	RGDALTS	1915	1.45
614	TTGGCTGGGC	1610	RSDHLAR	1712	QSSDLQR	1814	RSDSLTK	1916	2
616	GAGGAGGATG	1611	QSSNLQR	1713	RSDNLAR	1815	RSDNLQR	1917	0.08
617	AAGGGGGGG	1612	RSDHLSR	1714	RSDHLTR	1816	RKDNMTA	1918	1
618	AAGGGGGGG	1613	RSDHLSR	1715	RSDHLTR	1817	RKDNMTQ	1919	0.55
619	AAGGGGGGG	1614	RSDHLSR	1716	RSDHLTR	1818	RKDNMTN	1920	1.34
620	AAGGGGGGG	1615	RSDHLSR	1717	RSDHLTR	1819	RLDNRTA	1921	0.54
621	AAGGGGGGG	1616	RSDHLSR	1718	RSDHLTR	1820	RLDNRTQ	1922	0.75
624	ACGGATGTCT	1617	DRSALAR	1719	TSANLAR	1821	RSDTLRS	1923	7
628	TTGTAGGGGA	1618	RSDHLTR	1720	RSDNLTT	1822	RGDALTS	1924	130
629	TTGTAGGGGA	1619	RSSHLTR	1721	RSDNLTT	1823	RGDALTS	1925	150
630	CGGGGAGAGT	1620	RSDNLAR	1722	QSGHLQR	1824	RSDHLRE	1926	37.5
646	TTGGTGGAAG	1621	QSGNLAR	1723	RSDALAR	1825	RGDALTS	1927	35
647	TTGGTGGAAG	1622	QSANLAR	1724	RSDALAR	1826	RGDALTS	1928	40
651	GTTGTGGAAT	1623	QSGNLSR	1725	RSDALAR	1827	NRATLAR	1929	67.5
652	TAGGAGGCTG	1624	QSSDLQR	1726	RSDNLAR	1828	RSDNLTT	1930	1.5
653	TAGGAGGCTG	1625	TTSDLTR	1727	RSDNLAR	1829	RSDNLTT	1931	5.5
654	TAGGCATAAA	1626	QSGNLRT	1728	QSGSLTR	1830	RSDNLTT	1932	105
655	TAGGCATAAA	1627	QSGNLRT	1729	QSSTLRR	1831	RSDNLTT	1933	1000
656	TAGGCATAAA	1628	QSGNLRT	1730	QSGSLTR	1832	RSDNLTS	1934	540
657	TAGGCATAAA	1629	QSGNLRT	1731	QSSTLRR	1833	RSDNLTS	1935	300
660	GAGGGAGTTC	1630	NRATLAR	1732	QSGHLTR	1834	RSDNLAR	1936	8.25
661	GAGGGAGTTC	1631	TTSALTR	1733	QSGHLTR	1835	RSDNLAR	1937	1.73
665	GCGGAGGCGC	1632	RSDDVTR	1734	RSDNLTR	1836	RSDDLRR	1938	12.5

689	AAGGCGGAGA	1633	RSDNLTR	1735	RSDELQR	1837	RLDNRTA	1939	82.5
692	AAGGCGGAGA	1634	RSDNLTR	1736	RSDELQR	1838	RSDNLQ	1940	51
693	AAGGCGGAGA	1635	RSDNLTR	1737	RADTLRR	1839	RLDNRTA	1941	95
694	AAGGCGGAGA	1636	RSDNLTR	1738	RADTLRR	1840	RSDNLQ	1942	28.5
695	GGGGCGGAGC	1637	RSSNLTR	1739	DRSHLAR	1841	RSDHLTR	1943	850
697	TGAGCGGCGG	1638	RSDELTR	1740	RSDELSR	1842	QSGHLTK	1944	200
698	TGAGCGGCGG	1639	RSDELTR	1741	RSDELSR	1843	QSHGLTS	1945	300
699	GCGGCGGCAG	1640	QSGSLTR	1742	RSDDLQR	1844	RSDEKR	1946	21.5
700	GCGGCGGCAG	1641	QSGDLTR	1743	RSDDLQR	1845	RSDEKR	1947	45
701	GCAGCGGAGC	1642	RSDNLAR	1744	RSDELQR	1846	QSGSLTR	1948	50.5
702	GCAGCGGAGC	1643	RSDNLAR	1745	RSDELQR	1847	QSGDLTR	1949	73.5
704	AAGGTGGCAG	1644	QSGDLTR	1746	RSDSLAR	1848	RSDNLQ	1950	5
705	GGGGTGGGGC	1645	RSDHLAR	1747	RSDSLAR	1849	RSDHLSR	1951	0.01
706	GGGGTGGGGC	1646	RSDHLAR	1748	RSDSLR	1850	RSDHLSR	1952	0.05
708	GAGTCGAA	1647	QSANLAR	1749	RQDTLVG	1851	RSDNLAR	1953	300
709	GAGTCGAA	1648	QSANLAR	1750	RKDVLVS	1852	RSDNLAR	1954	400
710	GAGTCGAA	1649	QSGNLAR	1751	RLDGLRT	1853	RSDNLAR	1955	400
711	GAGTCGAA	1650	QSGNLAR	1752	RQDTLVG	1854	RSDNLAR	1956	400
712	GGTGAGGAGT	1651	RSDNLAR	1753	RSDNLAR	1855	MSDHLSR	1957	9.5
713	GGTGAGGAGT	1652	RSDNLAR	1754	RSDNLAR	1856	MSHHLSR	1958	0.15
714	TGGGTCGCGG	1653	RSDELRR	1755	DRSALAR	1857	RSDHLTT	1959	200
715	TGGGTCGCGG	1654	RADTLRR	1756	DRSALAR	1858	RSDHLTT	1960	0.46
716	TTGGGAGCAC	1655	QSGSLTR	1757	QSGHLQR	1859	RGDALTS	1961	200
717	TTGGGAGCAC	1656	QSGSLTR	1758	QSGHLQR	1860	RSDALTK	1962	150
718	TTGGGAGCAC	1657	QSGSLTR	1759	QSGHLQR	1861	RSDALTR	1963	107.5
719	GGCATGGTGG	1658	RSDALTR	1760	RSDALTS	1862	DRSHLAR	1964	20
720	GAAGAGGATG	1659	TTSNLAR	1761	RSDNLAR	1863	QSGNLTR	1965	1.6
722	ATGGGGGTGG	1660	RSDALTR	1762	RSDHLTR	1864	RSDALRQ	1966	0.7
724	GGCATGGTGG	1661	RSDALTR	1763	RSDALRQ	1865	DRSHLAR	1967	2.5
725	GCTTGAGTTA	1662	QSSALAR	1764	QSGHLQK	1866	QSSDLQR	1968	3000
726	GAAGAGGATG	1663	QSSNLAR	1765	RSDNLAR	1867	QSGNLTR	1969	1.5
727	GCGGTGGCTC	1664	QSSDLTR	1766	RSDALSR	1868	RSDTLKK	1970	0.1
728	GGTGAGGAGT	1665	RSDNLAR	1767	RSDNLAR	1869	DSSKLSR	1971	15

729	GGAGGGGAGT 1666	RSDNLAR 1768	RSDHLSR 1870	QSGHLAR 1972	1000
730	TGGGTCGCGG 1667	RSDDLTR 1769	DRSALAR 1871	RSDHLTT 1973	1000
731	GTGGGGGAGA 1668	RSDNLAR 1770	RSDHLSR 1872	RSDALAR 1974	12
732	GCGGGTGGGG 1669	RSDHLAR 1771	QSSHLAR 1873	RSDDLTR 1975	22.5
733	GCGGGTGGGG 1670	RSDHLAR 1772	QSSHLAR 1874	RSDTLKK 1976	0.32
734	GGGGCTGGGT 1671	RSDHLAR 1773	QSSDLSR 1875	RSDHLSR 1977	0.25
735	GCGGTGGCTC 1672	QSSDLTR 1774	RSDALSR 1876	RSDERKR 1978	0.05
736	GAGGTGGGGA 1673	RSDHLAR 1775	RSDALSR 1877	RSDNLSR 1979	0.47
737	GGAGGGGAGT 1674	RSDNLAR 1776	RSDHLSR 1878	QRGHLR 1980	1000
740	AAGGTGGCAG 1675	QSGSLTR 1777	RSDALAR 1879	RSDNRTA 1981	12.5
741	AAGGCTGAGA 1676	RSDNLTR 1778	QSSDLQR 1880	RSDNLQ 1982	15
742	ACGGGGTTAT 1677	QRGALAS 1779	RSDHLSR 1881	RSDTLKQ 1983	29
743	ACGGGGTTAT 1678	QRGALAS 1780	RSDHLSR 1882	RSDTLQ 1984	10
744	ACGGGGTTAT 1679	QRSALAS 1781	RSDHLSR 1883	RSDTLKQ 1985	8.33
745	ACGGGGTTAT 1680	QRSALAS 1782	RSDHLSR 1884	RSDTLQ 1986	12.5
746	CTGGAAGCAT 1681	QSGSLTR 1783	QSGNLAR 1885	RSDALRE 1987	2.07
747	CTATTTTGGG 1682	RSDHLTT 1784	QSSALRT 1886	QSGALRE 1988	2000
748	TTGGACGGCG 1683	DSGHLTR 1785	DRSNLER 1887	RGDALTS 1989	112.3
749	TTGGACGGCG 1684	DRSHLTR 1786	DSSNLTR 1888	RGDALTS 1990	11.33
750	GAGGGAGCGA 1685	RSDELTR 1787	QSAHLAR 1889	RSDNLAR 1991	52
751	GGTGAGGAGT 1686	RSDNLAR 1788	RSDNLAR 1890	NRSHLAR 1992	7
752	GAGGTGGGGA 1687	RSHHLAR 1789	RSDALSR 1891	RSDNLSR 1993	31
757	CGGGCGGCTG 1688	QSSDLRR 1790	RSDELQR 1892	RSDHLRE 1994	14.5
758	CGGGCGGCTG 1689	QSSDLRR 1791	RADTLRR 1893	RSDHLRE 1995	16.5
759	TTGGACGGCG 1690	DSGHLTR 1792	DSSNLTR 1894	RGDALTS 1996	37
760	TTGGACGGCG 1691	DRSHLTR 1793	DRSNLER 1895	RGDALTS 1997	148.5
761	GCGGTGGCTC 1692	QSSDLQR 1794	RSDALSR 1896	RSDERKR 1998	6
762	GCGGTGGCTC 1693	QSSDLQR 1795	RSDALSR 1897	RSDTLKK 1999	18
776	ATGGACGGGT 1694	RSDHLAR 1796	DRSNLER 1898	RSDSLNQ 2000	0.4
777	ATGGACGGGT 1695	RSDHLAR 1797	DRSNLTR 1899	RSDALSA 2001	3.4
779	CGGGGAGCAG 1696	QSGSLTR 1798	QSGHLTR 1900	RSDHLAE 2002	0.5
780	CGGGGAGCAG 1697	QSGSLTR 1799	QSGHLTR 1901	RSDHLRA 2003	0.5
781	GGGGAGCAGC 1698	RSSNLRE 1800	RSDNLAR 1902	RSDHLTR 2004	4.25

783	TTGGGAGCGG 1699	RSEDLTR	1801	QSGHLQR 1903	RGDALTS 2005	2000
785	TTGGGAGCGG 1700	RSDTLKK	1802	QSGHLQR 1904	RSDALTS 2006	50
786	TTGGGAGCGG 1701	RSDTLKK	1803	QSGHLQR 1905	RGDALRS 2007	2000
787	AGGGAGGATG 1702	QSDNLAR	1804	RSDNLAR 1906	RSDHLTQ 2008	4
826	GAGGGAGCGA 1703	RSEDLTR	1805	QSGHLAR 1907	RSDNLAR 2009	2.75
827	GAGGGAGCGA 1704	RADTLRR	1806	QSGHLAR 1908	RSDNLAR 2010	1.2
882	GCGTGGGCGT 1705	RSEDLTR	1807	RSDHLTT 1909	RSDERKR 2011	0.01
883	GCGTGGGCGT 1706	RSEDLTR	1808	RSDHLTT 1910	RSDERKR 2012	1

TABLE 5

SBS#	TARGET	<u>SEQ</u> ID	<u>F1</u>	<u>SEQ</u> ID	<u>F2</u>	<u>SEQ</u> ID	<u>F3</u>	<u>SEQ</u> ID	<u>Kd</u> (nM)
903	ATGGAAGGG	2013	RSDHLAR	2513	QSGNLAR	3013	RSDALRQ	3513	1.027
904	AAGGGTGAC	2014	DSSNLTR	2514	QSSHLAR	3014	RSDNLTQ	3514	1
905	GTGGTGGTG	2015	RSSALTR	2515	RSDSLAR	3015	RSDSLAR	3515	1.15
908	AAGGTCTCA	2016	QSGDLRT	2516	DRSALAR	3016	RSDNLRQ	3516	50
909	GTGGAAGAA	2017	QSGNLSR	2517	QSGNLQR	3017	RSDALAR	3517	16.4
910	ATGGAAGAT	2018	QSSNLAR	2518	QSGNLQR	3018	RSDALAQ	3518	0.03
911	ATGGGTGCA	2019	QSGSLTR	2519	QSSHLAR	3019	RSDALAQ	3519	0.91
912	TCAGAGGTG	2020	RSDSLAR	2520	RSDNLTR	3020	QSGDLRT	3520	0.135
914	CAGGAAAAG	2021	RSDNLTQ	2521	QSGNLAR	3021	RSDNLRE	3521	1.26
915	CAGGAAAAG	2022	RSDNLRQ	2522	QSGNLAR	3022	RSDNLRE	3522	45.15
916	GAGGAAGGA	2023	QSGHLAR	2523	QSGNLAR	3023	RSDNLQR	3523	1.3
919	TCATAGTAG	2024	RSDNLTT	2524	RSDNLRT	3024	QSGDLRT	3524	250
920	GATGTGGTA	2025	QSSSLVR	2525	RSDSLAR	3025	TSANLSR	3525	4
921	AAGGTCTCA	2026	QSGDLRT	2526	DPGALVR	3026	RSDNLRQ	3526	11
922	AAGGTCTCA	2027	QSHDLTK	2527	DRSALAR	3027	RSDNLRQ	3527	4
923	AAGGTCTCA	2028	QSHDLTK	2528	DPGALVR	3028	RSDNLRQ	3528	2
926	GTGGTGGTG	2029	RSDALTR	2529	RSDSLAR	3029	RSDSLAR	3529	7.502
927	CAGGTTGAG	2030	RSDNLAR	2530	TSGSLTR	3030	RSDNLRE	3530	3.61
928	CAGGTTGAG	2031	RSDNLAR	2531	QSSALTR	3031	RSDNLRE	3531	25
929	CAGGTAGAT	2032	QSSNLAR	2532	QSATLAR	3032	RSDNLRE	3532	1.3
931	GAGGAAGAG	2033	RSDNLAR	2533	QSSNLVR	3033	RSDNLAR	3533	2
932	ATGGAAGGG	2034	RSDHLAR	2534	QSSNLVR	3034	RSDALRQ	3534	797
933	GACGAGGAA	2035	QSANLAR	2535	RSDNLAR	3035	DRSNLTR	3535	500
934	ATGGAAGAT	2036	QSSNLAR	2536	QSGNLQR	3036	RSDALTS	3536	0.07
935	ATGGGTGCA	2037	QSGSLTR	2537	QSSHLAR	3037	RSDALTS	3537	0.91
937	GTGGGGGCT	2038	QSSDLTR	2538	RSDHLTR	3038	RSDSLAR	3538	0.03
938	GTGGGGGCT	2039	QSSDLRR	2539	RSDHLTR	3039	RSDSLAR	3539	0.049
939	GGGGGCTGG	2040	RSDHLTT	2540	DRSHLAR	3040	RSDHLSK	3540	0.352

940	GGGGGCTGG	2041	RSDHLTK 2541	DRSHLAR 3041	RSDHLSK 3541	1.5
941	GGGGCTGGG	2042	RSDHLAR 2542	QSSDLRR 3042	RSDKLSR 3542	0.077
942	GGGGCTGGG	2043	RSDHLAR 2543	QSSDLRR 3043	RSDHLSK 3543	0.13
943	GGGGCTGGG	2044	RSDHLAR 2544	TSGELVR 3044	RSDKLSR 3544	0.067
944	GGGGCTGGG	2045	RSDHLAR 2545	TSGELVR 3045	RSDHLSK 3545	0.027
945	GGTGCGGTG	2046	RSDSLTR 2546	RADTLRR 3046	MSHHLSR 3546	0.027
946	GGTGCGGTG	2047	RSDSLTR 2547	RSDVLQR 3047	MSHHLSR 3547	0.027
947	GGTGCGGTG	2048	RSDSLTR 2548	RSEDLQR 3048	QSSHLAR 3548	0.013
948	GGTGCGGTG	2049	RSDSLTR 2549	RSDVLQR 3049	QSSHLAR 3549	0.017
962	GAGGCGGCA	2050	QSGSLTR 2550	RSEDLQR 3050	RSDNLAR 3550	0.015
963	GAGGCGGCA	2051	QSGSLTR 2551	RSDDLQR 3051	RSDNLAR 3551	0.015
964	GCGGCGGTG	2052	RSDALAR 2552	RSEDLQR 3052	RSDERKR 3552	0.041
965	GCGGCGGCC	2053	ERGDLTR 2553	RSEDLQR 3053	RSDERKR 3553	3.1
966	GAGGAGGCC	2054	ERGTLAR 2554	RSDNLSR 3054	RSDNLAR 3554	0.028
967	GAGGAGGCC	2055	DRSSLTR 2555	RSDNLSR 3055	RSDNLAR 3555	0.055
968	GAGGCCGCA	2056	QSGSLTR 2556	DRSSLTR 3056	RSDNLAR 3556	1.4
969	GAGGCCGCA	2057	QSGSLTR 2557	DRSDLTR 3057	RSDNLAR 3557	0.275
970	GTGGGCGCC	2058	ERGTLAR 2558	DRSHLAR 3058	RSDALAR 3558	1.859
971	GTGGGCGCC	2059	DRSSLTR 2559	DRSHLAR 3059	RSDALAR 3559	0.144
972	GTGGGCGCC	2060	ERGDLTR 2560	DRSHLAR 3060	RSDALAR 3560	1.748
973	GCCGCGGTC	2061	DRSALTR 2561	RSEDLQR 3061	ERGTLAR 3561	0.6
974	GCCGCGGTC	2062	DRSALTR 2562	RSEDLQR 3062	DRSDLTR 3562	0.038
975	CAGGCCGCT	2063	QSSDLTR 2563	DRSSLTR 3063	RSDNLRE 3563	1.1
976	CAGGCCGCT	2064	QSSDLTR 2564	DRSDLTR 3064	RSDNLRE 3564	4.12
977	CTGGCAGTG	2065	RSDSLTR 2565	QSGSLTR 3065	RSDALRE 3565	0.017
978	CTGGCAGTG	2066	RSDSLTR 2566	QSGDLTR 3066	RSDALRE 3566	1.576
979	CTGGCGGCG	2067	RSSDLTR 2567	RSEDLQR 3067	RSDALRE 3567	1.59
980	CTGGCGGCG	2068	RSDDLTR 2568	RSEDLQR 3068	RSDALRE 3568	2.2
981	CAGGCGGCG	2069	RSDDLTR 2569	RSEDLQR 3069	RSDNLRE 3569	0.375
982	CCGGGCTGG	2070	RSDHLTT 2570	DRSHLAR 3070	RSEDLRE 3570	0.03
983	CCGGGCTGG	2071	RSDHLTK 2571	DRSHLAR 3071	RSEDLRE 3571	1.385
984	GACGGCGAG	2072	RSDNLAR 2572	DRSHLAR 3072	DRSNLTR 3572	1.6
985	GACGGCGAG	2073	RSDNLAR 2573	DRSHLAR 3073	EKANLTR 3573	0.965

986	GGTGCTGAT	2074	QSSNLQR 2574	QSSDLQR 3074	MSHHLSR 3574	1.6
987	GGTGCTGAT	2075	QSSNLQR 2575	QSSDLQR 3075	TSGHLVR 3575	33.55
988	GGTGCTGAT	2076	TSGNLVR 2576	QSSDLQR 3076	MSHHLSR 3576	0.15
989	GGTGAGGGG	2077	RSDHLAR 2577	RSDNLAR 3077	MSHHLSR 3577	1.9
990	AAGGTGGGC	2078	DRSHLTR 2578	RSDSLAR 3078	RSDNLTQ 3578	5.35
991	AAGGTGGGC	2079	DRSHLTR 2579	SSGSLVR 3079	RSDNLTQ 3579	0.06
993	GGGGCTGGG	2080	RSDHLAR 2580	TSSELVR 3080	RSDHLSR 3580	3.1
994	GGGGGCTGG	2081	RSDHLTK 2581	DRSHLAR 3081	RSDHLSR 3581	0.03
995	GGGGAGGAA	2082	QSANLAR 2582	RSDNLAR 3082	RSDHLSK 3582	0.08
996	CAGTTGGTC	2083	DRSALAR 2583	RSDALTS 3083	RSDNLRE 3583	9.6
997	AGAGAGGCT	2084	QSSDLTR 2584	RSDNLAR 3084	QSGHLNQ 3584	1.65
998	ACGTAGTAG	2085	RSANLRT 2585	RSDNLTK 3085	RSDTLKQ 3585	0.23
999	AGAGAGGCT	2086	QSSDLTR 2586	RSDNLAR 3086	QSGKLTQ 3586	0.6
1000	CAGTTGGTC	2087	DRSALAR 2587	RSDALTR 3087	RSDNLRE 3587	11.15
1001	GGAGCTGAC	2088	EKANLTR 2588	QSSDLSR 3088	QRAHLAR 3588	1.8
1002	GCGGAGGAG	2089	RSDNLVR 2589	RSDNLAR 3089	RSDERKR 3589	0.028
1003	ACGTAGTAG	2090	RSANLRT 2590	RSDNLTK 3090	RSDTLRS 3590	0.118
1004	ACGTAGTAG	2091	RSDNLTT 2591	RSDNLTK 3091	RSDTLRS 3591	1.4
1006	GTAGGGGCG	2092	RSDDLTR 2592	RSDHLTR 3092	QRASLTR 3592	0.898
1007	GAGAGAGAT	2093	QSSNLQR 2593	QSGHLTR 3093	RLHNLAR 3593	167
1008	GAGATGGAG	2094	RSDNLSR 2594	RSDSLTQ 3094	RLHNLAR 3594	0.4
1009	GAGATGGAG	2095	RSDNLSR 2595	RSDSLTQ 3095	RSDNLSR 3595	1.9
1010	GAGAGAGAT	2096	QSSNLQR 2596	QSGHLTR 3096	RSDNLAR 3596	8.2
1011	TTGGTGGCG	2097	RSADLTR 2597	RSDSLAR 3097	RSDSLTK 3597	0.03
1012	GACGTAGGG	2098	RSDHLTR 2598	QSSSLVR 3098	DRSNLTR 3598	0.032
1013	GAGAGAGAT	2099	QSSNLQR 2599	QSGHLNQ 3099	RSDNLAR 3599	0.15
1014	GACGTAGGG	2100	RSDHLTR 2600	QSGSLTR 3100	DRSNLTR 3600	0.01
1015	GCGGAGGAG	2101	RSDNLVR 2601	RSDNLAR 3101	RSDTLKK 3601	0.008
1016	CAGTTGGTC	2102	DRSALAR 2602	RSDSLTK 3102	RSDNLRE 3602	0.09
1017	CTGGATGAC	2103	EKANLTR 2603	TSGNLVR 3103	RSDALRE 3603	0.233
1018	GTAGTAGAA	2104	QSANLAR 2604	QSSSLVR 3104	QRASLAR 3604	7.2
1019	AGGGAGGAG	2105	RSDNLAR 2605	RSDNLAR 3105	RSDHLTQ 3605	0.022
1020	ACGTAGTAG	2106	RSDNLTT 2606	RSDNLTK 3106	RSDTLKQ 3606	0.69

1022	GAGGAGGTG	2107	RSDALAR	2607	RSDNLAR	3107	RSDNLAR	3607	0.01
1024	GGGGAGGAA	2108	QSANLAR	2608	RSDNLAR	3108	RSDHLSR	3608	0.08
1025	GAGGAGGTG	2109	QSSALTR	2609	QSSSLVR	3109	RSDTLTQ	3609	0.115
1026	GTGGCTTGT	2110	MSHHLKE	2610	QSSDLR	3110	RSDALAR	3610	0.076
1027	GCGGCGGTG	2111	RSDALAR	2611	RSDELQR	3111	RSDELQR	3611	0.054
1032	GGTGCTGAT	2112	TSGNLVR	2612	QSSDLQR	3112	TSGHLVR	3612	0.52
1033	GTGTTCTGT	2113	RSDALAR	2613	DRSALTT	3113	RSDALAR	3613	685.2
1034	GTGTTCTGT	2114	RSDALAR	2614	DRSALTK	3114	RSDALAR	3614	14.55
1035	GTGTTCTGT	2115	RSDALAR	2615	DRSALRT	3115	RSDALAR	3615	56
1037	GTAGGGGCA	2116	QSGSLTR	2616	RSDHLSR	3116	QRASLAR	3616	0.05
1038	GTAGGGGCA	2117	QTGELRR	2617	RSDHLSR	3117	QRASLAR	3617	0.152
1039	GGGGCTGGG	2118	RSDHLSR	2618	TSGELVR	3118	RSDHLTR	3618	1.37
1040	GGGGCTGGG	2119	RSDHLSR	2619	QSSDLQR	3119	RSDHLSK	3619	0.05
1041	TCATAGTAG	2120	RSDNLTT	2620	RSDNLRT	3120	QSHDLTK	3620	2.06
1043	CAGGGAGAG	2121	RSDNLAR	2621	QSGHLTR	3121	RSDNLRE	3621	0.16
1044	CAGGGAGAG	2122	RSDNLAR	2622	QRAHLER	3122	RSDNLRE	3622	1.07
1045	GGGGCAGGA	2123	QSGHLAR	2623	QSGSLTR	3123	RSDHLSR	3623	0.15
1046	GGGGCAGGA	2124	QSGHLAR	2624	QSGDLRR	3124	RSDHLSR	3624	0.09
1047	GGGGCAGGA	2125	QRAHLER	2625	QSGSLTR	3125	RSDHLSR	3625	24.7
1048	CAGGCTGTA	2126	QSGALTR	2626	QSSDLQR	3126	RSDNLRE	3626	1.387
1049	CAGGCTGTA	2127	QRASLAR	2627	QSSDLQR	3127	RSDNLRE	3627	55.6
1050	CAGGCTGTA	2128	QSSSLVR	2628	QSSDLQR	3128	RSDNLRE	3628	0.125
1051	GAGGCTGAG	2129	RSDNLTR	2629	QSSDLQR	3129	RSDNLVR	3629	0.02
1052	TAGGACGGG	2130	RSDHLAR	2630	EKANLTR	3130	RSDNLTT	3630	0.28
1053	TAGGACGGG	2131	RSDHLAR	2631	DRSNLTR	3131	RSDNLTT	3631	0.025
1054	GCTGCAGGG	2132	RSDHLAR	2632	QSGSLTR	3132	QSSDLQR	3632	0.033
1055	GCTGCAGGG	2133	RSDHLAR	2633	QSGSLTR	3133	TSGDLTR	3633	18.73
1056	GCTGCAGGG	2134	RSDHLAR	2634	QSGSLTR	3134	QSSDLQR	3634	0.045
1057	GCTGCAGGG	2135	RSDHLAR	2635	QSGDLTR	3135	TSGDLTR	3635	0.483
1058	GGGGCCGCG	2136	RSDELTR	2636	DRSSLTR	3136	RSDHLSR	3636	6.277
1059	GGGGCCGCG	2137	RSDELTR	2637	DRSDLTR	3137	RSDHLSR	3637	0.152
1060	GCGGAGGCC	2138	ERGTLAR	2638	RSDNLAR	3138	RSDEKRR	3638	0.69
1061	GTTGCGGGG	2139	RSDHLAR	2639	RSDELQR	3139	QSSALTR	3639	0.165

1062	GTTGCGGGG	2140	RSDHLAR	2640	RSEDLQR	3140	TSGSLTR	3640	0.068
1063	GTTGCGGGG	2141	RSDHLAR	2641	RSEDLQR	3141	MSHALSR	3641	0.96
1064	GCGGCAGTG	2142	RSDALTR	2642	QSGSLTR	3142	RSDERKR	3642	0.453
1065	TGGGGCGGG	2143	RSDHLAR	2643	DRSHLAR	3143	RSDHLTT	3643	1.37
1066	GAGGGCGGT	2144	QSSHLTR	2644	DRSHLAR	3144	RSDNLVR	3644	0.15
1067	GAGGGCGGT	2145	TSGHLVR	2645	DRSHLAR	3145	RSDNLVR	3645	1.37
1068	GCAGGGGGC	2146	DRSHLTR	2646	RSDHLTR	3146	QSGDLTR	3646	2.05
1069	GCAGGCGGT	2147	DRSHLTR	2647	RSDHLTR	3147	QSGSLTR	3647	0.1
1070	GGGGCAGGC	2148	DRSHLTR	2648	QSGSLTR	3148	RSDHLSR	3648	0.456
1071	GGGGCAGGC	2149	DRSHLTR	2649	QSGDLTR	3149	RSDHLSR	3649	0.2
1072	GGATTGGCT	2150	QSSDLTR	2650	RSDALTT	3150	QRAHLAR	3650	0.46
1073	GGATTGGCT	2151	QSSDLTR	2651	RSDALTK	3151	QRAHLAR	3651	1.37
1075	GTGTTGGCG	2152	RSEDLTR	2652	RSDALTK	3152	RSDALTR	3652	0.915
1076	GCGGCAGCG	2153	RSEDLTR	2653	QSGSLTR	3153	RSDERKR	3653	4.1
1077	GCGGCAGCG	2154	RSEDLTR	2654	QSGDLRR	3154	RSDERKR	3654	6.2
1078	GGGGGGGCC	2155	ERGTLAR	2655	RSDHLSR	3155	RSDHLSR	3655	0.2
1079	GGGGGGGCC	2156	ERGDLTR	2656	RSDHLSR	3156	RSDHLSR	3656	4.1
1080	CTGGAGGCG	2157	RSEDLTR	2657	RSDNLAR	3157	RSDALRE	3657	1.37
1081	GGGGAGGTG	2158	RSDALTR	2658	RSDNLTR	3158	RSDHLSR	3658	0.05
1082	CTGGCGGCG	2159	RSEDLTR	2659	RSEDLTR	3159	RSDALRE	3659	0.152
1083	CTGGTGGCA	2160	QSGDLTR	2660	RSDALSR	3160	RSDALRE	3660	0.152
1084	GGTGAGGCG	2161	RSEDLTR	2661	RSDNLAR	3161	MSHHLSR	3661	0.5
1085	GGTGAGGCG	2162	RSEDLTR	2662	RSDNLAR	3162	QSSHLAR	3662	0.46
1086	GGGGCTGGG	2163	RSDHLSR	2663	QSSDLQR	3163	RSDHLTR	3663	0.1
1087	CGGGCGGCC	2164	ERGDLTR	2664	RSEDLQR	3164	RSDHLAE	3664	1.24
1088	CGGGCGGCC	2165	ERGDLTR	2665	RSEDLQR	3165	RSDHLRE	3665	0.905
1089	GACGAGGCT	2166	QSSDLRR	2666	RSDNLAR	3166	DRSNLTR	3666	0.171
1090	AAGGCGCTG	2167	RSDALRE	2667	RSEDLQR	3167	RSDNLTQ	3667	30.3
1091	GTAGAGGAC	2168	DRSNLTR	2668	RSDNLAR	3168	QRASLAR	3668	0.085
1092	GCCTTGGCT	2169	QSSDLRR	2669	RGDALTS	3169	DRSDLTR	3669	2.735
1093	GCGGAGTCG	2170	RSADLRT	2670	RSDNLAR	3170	RSDERKR	3670	0.046
1094	GCGGTTGGT	2171	TSGHLVR	2671	QSSALTR	3171	RSDERKR	3671	12.34
1095	GGGGGAGCC	2172	ERGDLTR	2672	QRAHLER	3172	RSDHLSR	3672	0.395

1096	GGGGGAGCC	2173	DRSSLTR 2673	QRAHLER 3173	RSDHLSR 3673	0.019
1097	GAGGCCGAA	2174	QSANLAR 2674	DCRDLAR 3174	RSDNLAR 3674	0.77
1098	GCCGGGGAG	2175	RSDNLTR 2675	RSDHLTR 3175	DRSDLTR 3675	0.055
1099	GCGGAGTCG	2176	TSGHLVR 2676	TSGSLTR 3176	RSDERKR 3676	0.45
1100	GTGTTGGTA	2177	QSGALTR 2677	RGDALTS 3177	RSDALTR 3677	1.4
1101	ATGGGAGTT	2178	TTSALTR 2678	QRAHLER 3178	RSDALRQ 3678	0.065
1102	AAGGCAGAA	2179	QSANLAR 2679	QSGSLTR 3179	RSDNLTQ 3679	8.15
1103	AAGGCAGAA	2180	QSANLAR 2680	QSGDLTR 3180	RSDNLTQ 3680	1.4
1104	CGGGCAGCT	2181	QSSDLRR 2681	QSGSLTR 3181	RSDHLRE 3681	0.08
1105	CTGGCAGCC	2182	ERGDLTR 2682	QSGDLTR 3182	RSDALRE 3682	2.45
1106	CTGGCAGCC	2183	DRSSLTR 2683	QSGDLTR 3183	RSDALRE 3683	0.19
1107	GCGGGAGTT	2184	QSSALAR 2684	QRAHLER 3184	RSDERKR 3684	0.06
1108	CAGGCTGGA	2185	QSGHLAR 2685	TSGELVR 3185	RSDNLRE 3685	0.007
1109	AGGGGAGCC	2186	ERGDLTR 2686	QRAHLER 3186	RSDHLTQ 3686	0.347
1110	AGGGGAGCC	2187	DRSSLTR 2687	QRAHLER 3187	RSDHLTQ 3687	0.095
1111	CTGGTAGGG	2188	RSDHLAR 2688	QSSSLVR 3188	RSDALRE 3688	0.095
1112	CTGGTAGGG	2189	RSDHLAR 2689	QSATLAR 3189	RSDALRE 3689	0.125
1113	CTGGGGGCA	2190	QSGDLTR 2690	RSDHLTR 3190	RSDALRE 3690	0.06
1114	CAGGTTGAT	2191	QSSNLAR 2691	TSGSLTR 3191	RSDNLRE 3691	2.75
1115	CAGGTTGAT	2192	QSSNLAR 2692	QSSALTR 3192	RSDNLRE 3692	0.7
1116	CCGGAAGCG	2193	RSDELTR 2693	QSSNLVR 3193	RSDELRE 3693	12.3
1117	GCAGCGCAG	2194	RSSNLRE 2694	RSDELTR 3194	QSGSLTR 3694	2.85
1118	TAGGGAGTC	2195	DRSALTR 2695	QRAHLER 3195	RSDNLTT 3695	1.4
1119	TGGGAGGGT	2196	TSGHLVR 2696	RSDNLAR 3196	RSDHLTT 3696	0.1
1120	AGGGACGCG	2197	RSDELTR 2697	DRSNLTR 3197	RSDHLTQ 3697	2.735
1121	CTGGTGGCC	2198	ERGDLTR 2698	RSDALTR 3198	RSDALRE 3698	2.76
1122	CTGGTGGCC	2199	DRSSLTR 2699	RSDALTR 3199	RSDALRE 3699	0.101
1123	TAGGAAGCA	2200	QSGSLTR 2700	QSGNLAR 3200	RSDNLTT 3700	0.065
1124	GTGGATGGA	2201	QSGHLAR 2701	TSGNLVR 3201	RSDALTR 3701	0.101
1126	TTGGCTATG	2202	RSDALTS 2702	TSGELVR 3202	RGDALTS 3702	0.46
1127	CAGGGGGTT	2203	QSSALAR 2703	RSDHLTR 3203	RSDNLRE 3703	0.1
1128	AAGGTCGCC	2204	ERGDLTR 2704	DPGALVR 3204	RSDNLTQ 3704	5.45
1130	GGTGCAGAC	2205	DRSNLTR 2705	QSGDLTR 3205	MSHHLSR 3705	0.1

1131	GTGGGAGCC	2206	ERGDLTR 2706	QRAHLER 3206	RSDALTR 3706	0.95
1132	GGGGCTGGA	2207	QSGHLAR 2707	TSGELVR 3207	RSDHLSR 3707	0.055
1133	GGGGCTGGA	2208	QRAHLER 2708	TSGELVR 3208	RSDHLSR 3708	0.5
1134	TGGGGGTGG	2209	RSDHLTT 2709	RSDHLTR 3209	RSDHLTT 3709	0.067
1135	GCGGCGGGG	2210	RSDHLAR 2710	RSEDLQR 3210	RSDERKR 3710	0.025
1136	CCGGGAGTG	2211	RSDALTR 2711	QRAHLER 3211	RSDTLRE 3711	0.225
1137	CCGGGAGTG	2212	RSSALTR 2712	QRAHLER 3212	RSDTLRE 3712	0.085
1138	CAGGGGGTA	2213	QSGALTR 2713	RSDHLTR 3213	RSDNLRE 3713	0.027
1139	ACGGCCGAG	2214	RSDNLAR 2714	DRSDLTR 3214	RSDTLTQ 3714	0.535
1140	AAGGGTGCG	2215	RSEDLTR 2715	QSSHLAR 3215	RSDNLTQ 3715	0.3
1141	ATGGACTTG	2216	RGDALTS 2716	DRSNLTR 3216	RSDALTQ 3716	1.7
1148	TTGGAGGAG	2217	RSDNLTR 2717	RSDNLTR 3217	RGDALTS 3717	0.006
1149	TTGGAGGAG	2218	RSDNLTR 2718	RSDNLTR 3218	RSDALTK 3718	0.004
1150	GAAGAGGCA	2219	QSGSLTR 2719	RSDNLTR 3219	QSGNLTR 3719	0.004
1151	GTAGTATGG	2220	RSDHLTT 2720	QRSALAR 3220	QRASLAR 3720	1.63
1152	AAGGCTGGA	2221	QSGHLAR 2721	TSGELVR 3221	RSDNLTQ 3721	1.605
1153	AAGGCTGGA	2222	QRAHLAR 2722	TSGELVR 3222	RSDNLTQ 3722	8.2
1154	CTGGCGTAG	2223	RSDNLTT 2723	RSEDLQR 3223	RSDALRE 3723	1.04
1156	ATGGTTGAA	2224	QSANLAR 2724	QSSALTR 3224	RSDALRQ 3724	7.2
1157	ATGGTTGAA	2225	QSANLAR 2725	TSGSLTR 3225	RSDALRQ 3725	0.885
1158	AGGGGAGAA	2226	QSANLAR 2726	QSGHLTR 3226	RSDHLTQ 3726	0.1
1159	AGGGGAGAA	2227	QSANLAR 2727	QRAHLER 3227	RSDHLTQ 3727	0.555
1160	TGGGAAGGC	2228	DRSHLAR 2728	QSSNLVR 3228	RSDHLTT 3728	0.415
1161	GAGGCCGGC	2229	DRSHLAR 2729	DRSDLTR 3229	RSDNLAR 3729	0.45
1162	GTGTTGGTA	2230	QSGALTR 2730	RADALMV 3230	RSDALTR 3730	0.465
1163	GTGTGAGCC	2231	ERGDLTR 2731	QSGHLTT 3231	RSDALTR 3731	1.45
1164	GTGTGAGCC	2232	ERGDLTR 2732	QSVHLQS 3232	RSDALTR 3732	15.4
1165	GCGAAGGTG	2233	RSDALTR 2733	RSDNLTQ 3233	RSDERKR 3733	1.4
1166	GCGAAGGTG	2234	RSDALTR 2734	RSDNLTQ 3234	RSSDRKR 3734	0.195
1167	GCGAAGGTG	2235	RSDALTR 2735	RSDNLTQ 3235	RSHDRKR 3735	0.95
1168	AAGGCGCTG	2236	RSDALRE 2736	RSSDLTR 3236	RSDNLTQ 3736	2.8
1169	GTAGAGGAC	2237	DRSNLTR 2737	RSDNLAR 3237	QSSSLVR 3737	0.053
1170	GCCTTGGCT	2238	QSSDLRR 2738	RADALMV 3238	DRSDLTR 3738	2.75

1171	GCGGAGTCG	2239	RSDDLRT 2739	RSDNLAR 3239	RSDEKRR 3739	0.18
1172	GCCGGGGAG	2240	RSDNLTR 2740	RSDHLTR 3240	ERGDLTR 3740	0.01
1173	GCTGAAGGG	2241	RSDHLSR 2741	QSGNLAR 3241	QSSDLRR 3741	0.008
1174	GCTGAAGGG	2242	RSDHLSR 2742	QSSNLVR 3242	QSSDLRR 3742	0.018
1175	AAGGTCGCC	2243	DRSDLTR 2743	DPGALVR 3243	RSDNLTQ 3743	8.9
1176	GTGGGAGCC	2244	DRSDLTR 2744	QRAHLER 3244	RSDALTR 3744	4.1
1177	CCGGGCGCA	2245	QSGSLTR 2745	DRSHLAR 3245	RSDTLRE 3745	4.1
1178	GAGGATGGC	2246	DRSHLAR 2746	TSGNLVR 3246	RSDNLAR 3746	0.085
1179	GCAGCGCAG	2247	RSSNLRE 2747	RSSDLTR 3247	QSGSLTR 3747	2.735
1180	AAGGAAAGA	2248	QSGHLNQ 2748	QSGNLAR 3248	RSDNLTQ 3748	4.825
1181	TTGGCTATG	2249	RSDALRQ 2749	TSGELVR 3249	RGDALTS 3749	8.2
1182	CAGGAAGGC	2250	DRSHLAR 2750	QSGNLAR 3250	RSDNLRE 3750	1.48
1183	CAGGAAGGC	2251	DRSHLAR 2751	QSSNLVR 3251	RSDNLRE 3751	1.935
1184	AAGGAAAGA	2252	KNWKLQA 2752	QSGNLAR 3252	RSDNLTQ 3752	2.785
1185	AAGGAAAGA	2253	KNWKLQA 2753	QSHNLAR 3253	RSDNLTQ 3753	5.25
1186	GCCGAGGTG	2254	RSDSLLR 2754	RSKNLQR 3254	ERGTLAR 3754	27.5
1187	CTGGTGGGC	2255	DRSHLAR 2755	RSDALTR 3255	RSDALRE 3755	0.006
1188	GTAGTATGG	2256	RSDHLTT 2756	QSSSLVR 3256	QRASLAR 3756	2.74
1189	ATGGTTGAA	2257	QSANLAR 2757	TSGALTR 3257	RSDALRQ 3757	1.51
1190	ATGGCAGTG	2258	RSDALTR 2758	QSGDLTR 3258	RSDSLNQ 3758	1.484
1191	ATGGCAGTG	2259	RSDALTR 2759	QSGSLTR 3259	RSDSLNQ 3759	5.325
1192	ATGGCAGTG	2260	RSDALTR 2760	QSGDLTR 3260	RSDALTQ 3760	2.364
1193	ATGGCAGTG	2261	RSDALTR 2761	QSGSLTR 3261	RSDALTQ 3761	3.125
1194	GAGAAGGTG	2262	RSDALTR 2762	RSDNRTA 3262	RSDNLTR 3762	2.19
1195	GAGAAGGTG	2263	RSDALTR 2763	RSDNRTA 3263	RSSNLTR 3763	2.8
1197	GAAGGTGCC	2264	ERGDLTR 2764	MSHHLSR 3264	QSGNLTR 3764	14.8
1199	ATGGAGAAG	2265	RSDNRTA 2765	RSDNLTR 3265	RSDALTQ 3765	3.428
1200	ATGGAGAAG	2266	RSDNRTA 2766	RSSNLTR 3266	RSDALTQ 3766	16.87
1201	ATGGAGAAG	2267	RSDNRTA 2767	RSHNLTR 3267	RSDALTQ 3767	14.8
1202	CTGGAGTAC	2268	DRSNLRT 2768	RSDNLTR 3268	RSDALRE 3768	2.834
1203	GGAGTACTG	2269	RSDALRE 2769	QRSALAR 3269	QRAHLAR 3769	2.945
1204	GGAGTACTG	2270	RSDALRE 2770	QSSSLVR 3270	QRAHLAR 3770	4.38
1205	CGGGCAGCT	2271	QSSDLRR 2771	QSGDLTR 3271	RSDHLRE 3771	0.9

1206	GCGGGAGTT	2272	TTSALTR 2772	QRAHLER 3272	RSDERKR 3772	0.034
1207	CAGGCTGGA	2273	QRAHLER 2773	TSGELVR 3273	RSDNLRE 3773	0.45
1209	CCGGAAGCG	2274	RSDELTR 2774	QSSNLVR 3274	RSDTLRE 3774	19.28
1211	GCAGCGCAG	2275	RSDNLRE 2775	RSDELTR 3275	QSGSLTR 3775	6.5
1212	CAGGGGGTT	2276	TTSALTR 2776	RSDHLTR 3276	RSDNLRE 3776	0.05
1213	GAAGAAGAG	2277	RSDNLTR 2777	QSSNLVR 3277	QSGNLTR 3777	12.3
1214	ATGGGAGTT	2278	TTSALTR 2778	QRAHLER 3278	RSDALTQ 3778	0.46
1215	GTGGGGGCT	2279	QSSDLRR 2779	RSDHLTR 3279	RSDALTR 3779	0.003
1217	GAAGAGGCA	2280	QSGSLTR 2780	RSDNLTR 3280	QSANLTR 3780	0.004
1218	GCGGTGAGG	2281	RSDHLTQ 2781	RSQALTR 3281	RSDERKR 3781	0.46
1219	AAGGAAAGG	2282	RSDHLTQ 2782	QSHNLAR 3282	RSDNLTQ 3782	0.68
1220	AAGGAAAGG	2283	RSDHLTQ 2783	QSGNLAR 3283	RSDNLTQ 3783	0.175
1221	AAGGAAAGG	2284	RSDHLTQ 2784	QSSNLVR 3284	RSDNLTQ 3784	1.4
1222	CAGGAGGGC	2285	DRSHLAR 2785	RSDNLAR 3285	RSDNLRE 3785	0.155
1223	ATGGACTTG	2286	RSDALTK 2786	DRSNLTR 3286	RSDALTQ 3786	7
1224	ATGGACTTG	2287	RADALMV 2787	DRSNLTR 3287	RSDALTQ 3787	12
1227	GAATAGGGG	2288	RSDHLSR 2788	RSDHLTK 3288	QSGNLAR 3788	25
1228	ACGGCCGAG	2289	RSDNLAR 2789	DRSDLTR 3289	RSDDLTTQ 3789	12
1229	AAGGGTGCG	2290	RSDELTR 2790	MSHLSR 3290	RSDNLTQ 3790	8.2
1230	AAGGGAGAC	2291	DRSNLTR 2791	QSGHLTR 3291	RSDNLTQ 3791	0.383
1231	AAGGGAGAC	2292	DRSNLTR 2792	QRAHLER 3292	RSDNLTQ 3792	0.213
1232	TGGGACCTG	2293	RSDALRE 2793	DRSNLTR 3293	RSDHLTT 3793	0.113
1233	TGGGACCTG	2294	RSDALRE 2794	DRSNLTR 3294	RSDHLTT 3794	0.635
1234	GAGTAGGCA	2295	QSGSLTR 2795	RSDNLTK 3295	RSDNLAR 3795	0.101
1236	GAGTAGGCA	2296	QSGSLTR 2796	RSDHLTT 3296	RSDNLAR 3796	0.065
1237	GAAGGAGAG	2297	RSDNLAR 2797	QRAHLER 3297	QSGNLAR 3797	0.065
1238	CTGGATGTT	2298	QSSALAR 2798	TSGNLVR 3298	RSDALRE 3798	0.313
1239	CAGGACGTG	2299	RSDALTR 2799	DPGNLVR 3299	RSDNLKD 3799	0.144
1240	GGGGAGGCA	2300	QSGSLTR 2800	RSDNLTR 3300	RSDHLSR 3800	0.056
1241	GAGGTGTCA	2301	QSHDLTK 2801	RSDALAR 3301	RSDNLAR 3801	0.027
1242	GGGGTTGAA	2302	QSANLAR 2802	TSGSLTR 3302	RSDHLSR 3802	0.02
1243	GGGGTTGAA	2303	QSANLAR 2803	QSSALTR 3303	RSDHLSR 3803	0.101
1244	GTCGCGGTG	2304	RSDALTR 2804	RSDELQR 3304	DRSALAR 3804	0.044

1245	GTGCGGGTG	2305	RSDALTR	2805	RSDELQR	3305	DSGSLTR	3805	0.102
1246	GTGGTTGCG	2306	RSDELTR	2806	TSGSLTR	3306	RSDALTR	3806	0.051
1247	GTGGTTGCG	2307	RSDELTR	2807	TSGALTR	3307	RSDALTR	3807	0.117
1248	GTCTAGGTA	2308	QSGALTR	2808	RSDNLTT	3308	DRSALAR	3808	5.14
1249	CCGGGAGCG	2309	RSDELTR	2809	QSGHLTR	3309	RSDTLRE	3809	0.26
1250	GAAGGAGAG	2310	RSDNLAR	2810	QSGHLTR	3310	QSGNLAR	3810	0.31
1252	CCGGCTGGA	2311	QRAHLER	2811	QSSDLTR	3311	RSDTLRE	3811	0.153
1253	CCGGGAGCG	2312	RSDELTR	2812	QRAHLER	3312	RSDTLRE	3812	0.228
1255	ACGTAGTAG	2313	RSDNLTT	2813	RSDNLTK	3313	RSDTLKQ	3813	0.69
1256	GGGGAGGAT	2314	QSSNLAR	2814	RSDNLQR	3314	RSDHLSR	3814	2
1257	GGGGAGGAT	2315	TTSNLAR	2815	RSDNLQR	3315	RSDHLSR	3815	1
1258	GGGGAGGAT	2316	QSSNLRR	2816	RSDNLQR	3316	RSDHLSR	3816	2
1259	GAGTGTGTG	2317	RSDSLLR	2817	DRDHLTR	3317	RSDNLAR	3817	1.5
1260	GAGTGTGTG	2318	RLDSLRL	2818	DRDHLTR	3318	RSDNLAR	3818	1.8
1261	TGCGGGGCA	2319	QSGDLTR	2819	RSDHLTR	3319	RRDTLHR	3819	0.2
1262	TGCGGGGCA	2320	QSGDLTR	2820	RSDHLTR	3320	RLDTLGR	3820	3
1263	TGCGGGGCA	2321	QSGDLTR	2821	RSDHLTR	3321	DSGHLAS	3821	21
1264	AAGTTGGTT	2322	TTSALTR	2822	RADALMV	3322	RSDNLQ	3822	0.21
1265	AAGTTGGTT	2323	TTSALTR	2823	RSDALTT	3323	RSDNLQ	3823	0.077
1266	CAGGGTGGC	2324	DRSHLTR	2824	QSSHLAR	3324	RSDNLRE	3824	6.1
1267	TAGGCAGTC	2325	DRSALTR	2825	QSGSLTR	3325	RSDNLTT	3825	6
1268	CTGTTGGCT	2326	QSSDLTR	2826	RADALMV	3326	RSDALRE	3826	1.52
1269	CTGTTGGCT	2327	QSSDLTR	2827	RSDALTT	3327	RSDALRE	3827	12.3
1270	TTGGATGGA	2328	QSGHLAR	2828	TSGNLVR	3328	RSDALTK	3828	0.4
1271	GTGGCACTG	2329	RSDALRE	2829	QSGSLTR	3329	RSDALTR	3829	0.915
1272	CAGGAGTCC	2330	DRSSLTT	2830	RSDNLAR	3330	RSDNLRE	3830	0.04
1273	CAGGAGTCC	2331	ERGDLT	2831	RSDNLAR	3331	RSDNLRE	3831	0.1
1274	GCATGGGAA	2332	QSANLSR	2832	RSDHLTT	3332	QSGSLTR	3832	0.306
1275	GCATGGGAA	2333	QRSNLVR	2833	RSDHLTT	3333	QSGSLTR	3833	0.326
1276	TAGGAAGAG	2334	RSDNLAR	2834	QRSNLVR	3334	RSDNLTT	3834	0.685
1277	GAAGAGGGG	2335	RSDHLAR	2835	RSDNLAR	3335	QSGNLTR	3835	0.421
1278	GAGTAGGCA	2336	QSGSLTR	2836	RSDNLRT	3336	RSDNLAR	3836	0.019
1279	GAGGTGTCA	2337	QSGDLRT	2837	RSDALAR	3337	RSDNLAR	3837	0.025

1282	TCGGTCGCC	2338	ERGDLTR	2838	DPGALVR	3338	RSDELRT	3838	74.1
1287	GTGGTAGGA	2339	QSGHLAR	2839	QSGALAR	3339	RSDALTR	3839	0.152
1288	CAGGGTGGC	2340	DRSHLTR	2840	QSSHLAR	3340	RSDNLTE	3840	4.1
1289	TAGGCAGTC	2341	DRSALTR	2841	QSGSLTR	3341	RSDNLTK	3841	1.37
1290	GTGGTGATA	2342	QSGALTQ	2842	RSHALTR	3342	RSDALTR	3842	24.05
1291	GTGGTGATA	2343	QQASLNA	2843	RSHALTR	3343	RSDALTR	3843	20.55
1292	TTGGATGGA	2344	QSGHLAR	2844	TSGNLVR	3344	RSDALTT	3844	4.12
1293	AAGGTAGGT	2345	TSGHLVR	2845	QSGALAR	3345	RSDNLTK	3845	0.457
1294	AAGGTAGGT	2346	MSHHLVR	2846	QSGALAR	3346	RSDNLTK	3846	2.75
1295	CAGGAGTCC	2347	DRSSLTT	2847	RSDNLAR	3347	RSDNLTE	3847	0.116
1296	CAGGAGTCC	2348	ERGDLT	2848	RSDNLAR	3348	RSDNLTE	3848	37
1297	TAGGAAGAG	2349	RSDNLAR	2849	QRSNLVR	3349	RSDNLTK	3849	0.05
1298	CAGGACGTG	2350	RSDLATR	2850	DPGNLVR	3350	RSDNLTE	3850	0.05
1300	GTCTAGGTA	2351	QSGALTR	2851	RSDNLTK	3351	DRSALAR	3851	0.46
1302	CCGGCTGGA	2352	QSGHLTR	2852	QSSDLTR	3352	RSDTLRE	3852	0.05
1303	TAGGAGTTT	2353	QRSALAS	2853	RSDNLAR	3353	RSDNLTT	3853	0.088
1306	CTGGCCTTG	2354	RSDALTT	2854	DCRDLAR	3354	RSDALRE	3854	2.285
1308	TGGGCAGCC	2355	ERGTLAR	2855	QSGSLTR	3355	RSDHLTT	3855	0.305
1309	TAGGAGTTT	2356	QSSALAS	2856	RSDNLAR	3356	RSDNLTT	3856	0.184
1310	TAGGAGTTT	2357	TTSALAS	2857	RSDNLAR	3357	RSDNLTT	3857	0.075
1311	TGGGCAGCC	2358	ERGDLAR	2858	QSGSLTR	3358	RSDHLTT	3858	0.91
1312	GGGGCGTGA	2359	QSGHLTK	2859	RSDELQR	3359	RSDHLSR	3859	0.23
1313	GGGGCGTGA	2360	QSGHLTT	2860	RSDELQR	3360	RSDHLSR	3860	0.09
1314	GTACAGTAG	2361	RSDNLTT	2861	RSDNLRE	3361	QSSSLVR	3861	3.09
1315	GTACAGTAG	2362	RSDNLTT	2862	RSDNLTE	3362	QSSSLVR	3862	9.27
1318	ATGGTGTGT	2363	TSSHLAS	2863	RSDALAR	3363	RSDALAQ	3863	0.048
1319	ATGGTGTGT	2364	MSHHLTT	2864	RSDALAR	3364	RSDALAQ	3864	0.228
1320	TTGGGAGAG	2365	RSDNLAR	2865	QRAHLER	3365	RSDALTT	3865	0.044
1321	TTGGGAGAG	2366	RSDNLAR	2866	QRAHLER	3366	RADALMV	3866	0.127
1322	GTGGGAATA	2367	QSGALTQ	2867	QSGHLTR	3367	RSDALTR	3867	0.799
1323	GTGGGAATA	2368	QLTGLNQ	2868	QSGHLTR	3368	RSDALTR	3868	0.744
1324	GTGGGAATA	2369	QQASLNA	2869	QSHHLTR	3369	RSDALTR	3869	18.52
1325	TTGGTTGGT	2370	TSGHLVR	2870	TSGSLTR	3370	RSDALTK	3870	0.306

1326	TTGGTTGGT	2371	TSGHLVR	2871	QSSALTR	3371	RSDALTK	3871	4.385
1327	TTGGTTGGT	2372	TSGHLVR	2872	TSGSLTR	3372	RSDALTT	3872	0.566
1328	TTGGTTGGT	2373	TSGHLVR	2873	QSSALTR	3373	RSDALTT	3873	7.95
1329	CTGGCCTGG	2374	RSDHLTT	2874	DRSDLTR	3374	RSDALRE	3874	0.68
1330	GAGGTGTGA	2375	QSGHLTT	2875	RSDALTR	3375	RSDNLAR	3875	0.175
1331	CTGGCCTGG	2376	RSDHLTT	2876	DCRDLAR	3376	RSDALRE	3876	0.388
1334	CCGGCGCTG	2377	RSDALRE	2877	RSSDLTR	3377	RSDDLRE	3877	0.31
1335	GACGCTGGC	2378	DRSHLTR	2878	QSSDLTR	3378	DSSNLTR	3878	1.4
1336	CGGGCTGGA	2379	QSGHLAR	2879	QSSDLTR	3379	RSDHLAE	3879	1.4
1337	CGGGCTGGA	2380	QSSHLAR	2880	QSSDLTR	3380	RSDHLAE	3880	0.235
1338	GGGATGGCG	2381	RSDELTR	2881	RSDALTQ	3381	RSDHLSR	3881	1.04
1339	GGGATGGCG	2382	RSDELTR	2882	RSDSLTQ	3382	RSDHLSR	3882	0.569
1340	GGGATGGCG	2383	RSDELTR	2883	RSDALTQ	3383	RSHHLSR	3883	0.751
1341	GGGATGGCG	2384	RSDELTR	2884	RSDSLTQ	3384	RSHHLSR	3884	4.1
1342	CAGGCGCAG	2385	RSDNLRE	2885	RSSDLTR	3385	RSDNLTE	3885	0.68
1343	CAGGCGCAG	2386	RSDNLTT	2886	RTSTLTR	3386	RSDNLTE	3886	37.04
1344	CCGGGCGAC	2387	DRSNLTR	2887	DRSHLAR	3387	RSDTLRE	3887	2.28
1346	GATGTGTGA	2388	QSGHLTT	2888	RSDALAR	3388	TSANLSR	3888	0.153
1347	CAGTGAATG	2389	RSDALTS	2889	QSHHLTT	3389	RSDNLTE	3889	8.23
1348	GGGTCACTG	2390	RSDALTA	2890	QAATLTT	3390	RSDHLSR	3890	2.58
1350	CAGTGAATG	2391	RSDALTQ	2891	QSGHLTT	3391	RSDNLTE	3891	74.1
1351	GGGTCACTG	2392	RSDALRE	2892	QSHDLTK	3392	RSDHLSR	3892	0.234
1352	GTGTGGGTC	2393	DRSALAR	2893	RSDHLTT	3393	RSDALTR	3893	0.023
1353	CTGGCGAGA	2394	QSGHLNQ	2894	RSDELQR	3394	RSDALRE	3894	56.53
1354	CTGGCGAGA	2395	KNWKLQA	2895	RSDELQR	3395	RSDALRE	3895	20.85
1355	GCTTTGGCA	2396	QSGSLTR	2896	RSDALTT	3396	QSSDLTR	3896	0.172
1356	GCTTTGGCA	2397	QSGSLTR	2897	RADALMV	3397	QSSDLTR	3897	0.034
1357	GACTTG GTA	2398	QSSSLVR	2898	RSDALTT	3398	DRSNLTR	3898	0.032
1358	GACTTG GTA	2399	QSSSLVR	2899	RADALMV	3399	DRSNLTR	3899	0.05
1360	CAGTTGTGA	2400	QSGHLTT	2900	RADALMV	3400	RSDNLTE	3900	41.7
1361	AAGGAAAAA	2401	QKTNLDT	2901	QSGNLQR	3401	RSDNLTQ	3901	0.835
1362	AAGGAAAAA	2402	QSGNLNQ	2902	QSGNLQR	3402	RSDNLTQ	3902	0.332
1363	AAGGAAAAA	2403	QKTNLDT	2903	QRSNLVR	3403	RSDNLTQ	3903	74.1

1364	ATGGGTGAA	2404	QSANLSR	2904	QSSHLAR	3404	RSDALAQ	3904	1.22
1365	ATGGGTGAA	2405	QRSNLVR	2905	QSSHLAR	3405	RSDALAQ	3905	0.152
1366	ATGGGTGAA	2406	QSANLSR	2906	TSGHLVR	3406	RSDALAQ	3906	22.63
1367	ATGGGTGAA	2407	QRSNLVR	2907	TSGHLVR	3407	RSDALAQ	3907	1.028
1368	CTGGGAGAT	2408	QSSNLAR	2908	QRAHLER	3408	RSDALRE	3908	0.051
1369	CTGGGAGAT	2409	QSSNLAR	2909	QSGHLTR	3409	RSDALRE	3909	0.227
1373	GTGGTGGGC	2410	DRSHLTR	2910	RSDALSR	3410	RSDALTR	3910	0.025
1374	CCGGCGGTG	2411	RSDALTR	2911	RSDELQR	3411	RSDELRE	3911	0.003
1375	CCGGCGGTG	2412	RSDALTR	2912	RSDDLQR	3412	RSDELRE	3912	0.008
1376	CCGGCGGTG	2413	RSDALTR	2913	RSDEKRR	3413	RSDELRE	3913	0.858
1377	CCGGCGGTG	2414	RSDALTR	2914	RSDELQR	3414	RSDDLRE	3914	0.012
1378	CCGGCGGTG	2415	RSDALTR	2915	RSDDLQR	3415	RSDDLRE	3915	0.012
1379	CCGGCGGTG	2416	RSDALTR	2916	RSDEKRR	3416	RSDDLRE	3916	0.25
1380	GCCGACGGT	2417	QSSHLTR	2917	DRSNLTR	3417	ERGDLTR	3917	0.076
1381	GCCGACGGT	2418	QSSHLTR	2918	DPGNLVR	3418	ERGDLTR	3918	0.23
1382	GCCGACGGT	2419	QSSHLTR	2919	DRSNLTR	3419	DCRDLAR	3919	3.1
1383	GCCGACGGT	2420	QSSHLTR	2920	DPGNLVR	3420	DCRDLAR	3920	1.74
1384	GGTGTGGGC	2421	DRSHLTR	2921	RSDALSR	3421	MSHHLSR	3921	0.013
1385	TGGGCAAGA	2422	QSGHLNQ	2922	QSGSLTR	3422	RSDHLTT	3922	0.229
1386	TGGGCAAGA	2423	ENWKLQA	2923	QSGSLTR	3423	RSDHLTT	3923	0.193
1389	CTGGCCTGG	2424	RSDHLTT	2924	DCRDLAR	3424	RSDALRE	3924	0.175
1393	TGGGAAGCT	2425	QSSDLRR	2925	QSGNLAR	3425	RSDHLTT	3925	0.1
1394	TGGGAAGCT	2426	QSSDLRR	2926	QSGNLAR	3426	RSDHLTK	3926	0.04
1395	GAAGAGGGA	2427	QSGHLQR	2927	RSDNLAR	3427	QSGNLAR	3927	0.025
1396	GAAGAGGGA	2428	QRAHLAR	2928	RSDNLAR	3428	QSGNLAR	3928	0.107
1397	GAAGAGGGA	2429	QSSHLAR	2929	RSDNLAR	3429	QSGNLAR	3929	0.14
1398	TAATGGGGG	2430	RSDHLSR	2930	RSDHLTT	3430	QSGNLRT	3930	0.065
1399	TGGGAGTGT	2431	TKQHLKT	2931	RSDNLAR	3431	RSDHLTT	3931	0.1
1400	CCGGGTGAG	2432	RSDNLAR	2932	QSSHLAR	3432	RSDDLRE	3932	0.371
1401	GAGTTGGCC	2433	ERGTLAR	2933	RADALMV	3433	RSDNLAR	3933	0.167
1402	CTGGAGTTG	2434	RGDALTS	2934	RSDNLAR	3434	RSDALRE	3934	0.15
1403	ATGGCAATG	2435	RSDALTQ	2935	QSGSLTR	3435	RSDALTQ	3935	0.07
1404	GAGGCAGGG	2436	RSDHLSR	2936	QSGSLTR	3436	RSDNLAR	3936	0.022

1405	GAGGCAGGG	2437	RSDHLSR 2937	QSGDLTR 3437	RSDNLAR 3937	0.045
1406	GAAGCGGAG	2438	RSDNLAR 2938	RSDELTR 3438	QSGNLAR 3938	0.025
1407	GCGGGCGCA	2439	QSGSLTR 2939	DRSHLAR 3439	RSDERKR 3939	0.585
1408	CCGGCAGGG	2440	RSDHLSR 2940	QSGSLTR 3440	RSDELRE 3940	0.305
1409	CCGGCAGGG	2441	RSDHLSR 2941	QSGSLTR 3441	RSDDLRE 3941	0.153
1410	CCGGCGGCG	2442	RSDELTR 2942	RSDELQR 3442	RSDELRE 3942	0.814
1411	TGAGGCGAG	2443	RSDNLAR 2943	DRSHLAR 3443	QSGHLTK 3943	0.282
1412	CTGGCCGTG	2444	RSDSLLR 2944	ERGTLAR 3444	RSDALRE 3944	0.172
1413	CTGGCCGCG	2445	RSDELTR 2945	DRSDLTR 3445	RSDALRE 3945	0.152
1414	CTGGCCGCG	2446	RSDELTR 2946	ERGTLAR 3446	RSDALRE 3946	0.914
1415	GCGGCCGAG	2447	RSDNLAR 2947	DRSDLTR 3447	RSDELQR 3947	0.102
1416	GCGGCCGAG	2448	RSDNLAR 2948	ERGTLAR 3448	RSDELQR 3948	0.153
1417	GAGTTGGCC	2449	ERGTLAR 2949	RGDALTS 3449	RSDNLAR 3949	1.397
1418	CTGGAGTTG	2450	RADALMV 2950	RSDNLAR 3450	RSDALRE 3950	0.241
1422	GGGTCGGCG	2451	RSDELTR 2951	RSDDLTT 3451	RSDHLSR 3951	0.064
1423	GGGTCGGCG	2452	RSDELTR 2952	RSDDLTK 3452	RSDHLSR 3952	0.034
1424	CAGGGCCCG	2453	RSDELRE 2953	DRSHLAR 3453	RSDNLRE 3953	1.37
1427	CAGGGCCCG	2454	RSDDLRE 2954	DRSHLAR 3454	RSDNLTE 3954	0.271
1428	TGAGGCGAG	2455	RSDNLAR 2955	DRSHLAR 3455	QSVHLQS 3955	0.102
1429	TGAGGCGAG	2456	RSDNLAR 2956	DRSHLAR 3456	QSGHLTT 3956	0.074
1430	TCGGCCGCC	2457	ERGTLAR 2957	DRSDLTR 3457	RSDDLTK 3957	0.352
1431	TCGGCCGCC	2458	ERGTLAR 2958	DRSDLTR 3458	RSDDLAS 3958	6.17
1432	TCGGCCGCC	2459	ERGTLAR 2959	ERGTLAR 3459	RSDDLTK 3959	1.778
1434	CTGGCCGTG	2460	RSDSLLR 2960	DRSDLTR 3460	RSDALRE 3960	0.051
1435	TAATGGGGG	2461	RSDHLSR 2961	RSDHLTT 3461	QSGNLTK 3961	0.057
1436	TGGGAGTGT	2462	TSDHLAS 2962	RSDNLAR 3462	RSDHLTT 3962	0.026
1439	GGAGTGTTA	2463	QRSALAS 2963	RSDALAR 3463	QSGHLQR 3963	0.075
1440	GGAGTGTTA	2464	QSGALTK 2964	RSDALAR 3464	QSGHLQR 3964	0.035
1441	ATAGCTGGG	2465	RSDHLSR 2965	QSSDLTR 3465	QSGALTQ 3965	0.262
1442	TGCTGGGCC	2466	ERGTLAR 2966	RSDHLTT 3466	DRSHLTK 3966	0.36
1443	TGGAAGGAA	2467	QSGNLAR 2967	RSDNLTK 3467	RSHHLTT 3967	0.22
1444	TGGAAGGAA	2468	QSGNLAR 2968	RSDNLTK 3468	RSHHLTT 3968	0.09
1445	TGGAAGGAA	2469	QSGNLAR 2969	RLDNLTA 3469	RSHHLTT 3969	0.182

1446	TGGAAGGAA	2470	QSGNLAR	2970	RLDNLTA	3470	RSSHLTT	3970	0.42
1454	GGAGAGGCT	2471	QSSDLRR	2971	RSDNLAR	3471	QSGHLQR	3971	0.01
1455	CGGGATGAA	2472	QSANLSR	2972	TSGNLVR	3472	RSDHLRE	3972	0.043
1456	GGAGAGGCT	2473	QSSDLRR	2973	RSDNLAR	3473	QRAHLAR	3973	0.016
1457	GCAGAGGAA	2474	QSANLSR	2974	RSDNLAR	3474	QSGSLTR	3974	0.014
1460	TTGGGGGAG	2475	RSDNLAR	2975	RSDHLTR	3475	RADALMV	3975	0.007
1461	GACGAGGAG	2476	RSANLAR	2976	RSDNLTR	3476	DRSNLTR	3976	0.014
1462	CGGGATGAA	2477	QSGNLAR	2977	TSGNLVR	3477	RSDHLRE	3977	0.05
1463	GAGGCTGTT	2478	TTSALTR	2978	QSSDLTR	3478	RSDNLAR	3978	0.003
1464	GACGAGGAG	2479	RSDNLAR	2979	RSDNLTR	3479	DRSNLTR	3979	0.002
1465	CTGGGAGTT	2480	TTSALTR	2980	QSGHLQR	3480	RSDALRE	3980	0.018
1466	CTGGGAGTT	2481	NRATLAR	2981	QSGHLQR	3481	RSDALRE	3981	0.017
1468	GGTGATGTC	2482	DRSALTR	2982	TSGNLVR	3482	MSHHLSR	3982	0.08
1469	GGTGATGTC	2483	DRSALTR	2983	TSGNLVR	3483	TSGHLVR	3983	0.28
1470	GGTGATGTC	2484	DRSALTR	2984	TSGNLVR	3484	QRAHLER	3984	0.156
1471	CTGGTTGGG	2485	RSDHLSR	2985	QSSALTR	3485	RSDALRE	3985	0.09
1472	TTGAAGGTT	2486	TTSALTR	2986	RSDNLTQ	3486	RADALMV	3986	3.22
1473	TTGAAGGTT	2487	TTSALTR	2987	RSDNLTQ	3487	RSDSLTT	3987	0.47
1474	TTGAAGGTT	2488	QSSALAR	2988	RSDNLTQ	3488	RADALMV	3988	1.39
1475	TTGAAGGTT	2489	QSSALAR	2989	RSDNLTQ	3489	RLHSLTT	3989	0.39
1476	TTGAAGGTT	2490	QSSALAR	2990	RSDNLTQ	3490	RSDSLTT	3990	0.305
1477	GCAGCCCGG	2491	RSDHLRE	2991	DRSDLTR	3491	QSGSLTR	3991	2.31
1479	GAAAGTTCA	2492	QSHDLTK	2992	MSHHLTQ	3492	QSGNLAR	3992	37.04
1480	GAAAGTTCA	2493	NKTDLGK	2993	TSGHLVQ	3493	QSGNLAR	3993	62.5
1481	GAAAGTTCA	2494	NKTDLGK	2994	TSDHLAS	3494	RSDELRE	3994	37.04
1482	CCGTGTGAC	2495	DRSNLTR	2995	TSDHLAS	3495	RSDELRE	3995	111.1
1483	CCGTGTGAC	2496	DRSNLTR	2996	MSHHLTT	3496	RSDELRE	3996	20.8
1484	GAAGTGGTA	2497	QSSSLVR	2997	RSDALSR	3497	QSGNLAR	3997	0.01
1485	AAGTGAGCT	2498	QSSDLRR	2998	QSGHLTT	3498	RSDNLTQ	3998	1.537
1486	GGGTTTGAC	2499	DRSNLTR	2999	TTSALAS	3499	RSDHLSR	3999	0.085
1487	TTGAAGGTT	2500	TTSALTR	3000	RSDNLTQ	3500	RLHSLTT	4000	0.188
1488	AAGTGGTAG	2501	QSSDLRR	3001	QSGHLTT	3501	RLDNRTQ	4001	5.64
1490	CTGGTTGGG	2502	RSDHLSR	3002	TSGSLTR	3502	RSDALRE	4002	0.04

1491	AAGGGTTCA	2503	NKTDLGK 3003	DSSKLSR 3503	RLDNRTA 4003	4.12
1492	AAGTGGTAG	2504	RSDNLTT 3004	RSDHLTT 3504	RSDNLTQ 4004	1.37
1493	AAGTGGTAG	2505	RSDNLTT 3005	RSDHLTT 3505	RLDNRTQ 4005	15.09
1494	GGGTTTGAC	2506	DRSNLTR 3006	QRSALAS 3506	RSDHLSR 4006	0.255
1496	TTGGGGGAG	2507	RSDNLAR 3007	RSDHLTR 3507	RSDALTT 4007	0.065
1497	GAGGCTCTT	2508	QSSALAR 3008	QSSDLTR 3508	RSDNLAR 4008	0.007
1498	GAGGTTGAT	2509	QSSNLAR 3009	QSSALTR 3509	RSDNLAR 4009	0.101
1499	GAGGTTGAT	2510	QSSNLAR 3010	TSGALTR 3510	RSDNLAR 4010	0.02
1500	GCAGAGGAA	2511	QSGNLAR 3011	RSDNLAR 3511	QSGSLTR 4011	0.003
1522	GCAATGGGT	2512	TSGHLVR 3012	RSDALTQ 3512	QSGDLTR 4012	0.08

TABLE 6

TRIPLET (5'→3')	FINGER (N → C)		
	F1	F2	F3
AGG			RXDHXXQ
ATG			RXDAXXQ
CGG			RXDHXXE
GAA		QXGNXXR	
GAC	DXSNXXR		DXSNXXR
GAG	RXDNXXR	RXSNXXR RXDNXXR	RXDNXXR
GAT	QXSNXXR TXSNXXR TXGNXXR	TXGNXXR	
GCA	QXGSXXR	QXGDXXR	
GCC	EXGTXXR		
GCG	RXDEXXR	RXDEXXR	RXDEXXR RXDTXXK
GCT	QXSDXXR	TXGEXXR QXSDXXR	
GGA		QXGHXXR	QXAHXXR
GGC	DXSHXXR	DXSHXXR	
GGG	RXDHXXR	RXDHXXR	RXDHXXR RXDHXXK
GGT			TXGHXXR
GTA		QXGSXXR QXATXXR	
GTG	RXDAXXR RXDSXXR	RXDAXXR	RXDAXXR
TAG		RXDNXXT	
TCG	RXDDXXK		
TGT		TXDHXXS	

WHAT IS CLAIMED IS:

1. A zinc finger which binds to a target subsite wherein amino acids –1 through +6 of the zinc finger and the nucleotide sequence of the target subsite are as specified in Table 6.

5 2. A zinc finger according to claim 1 wherein amino acids –1 through +6 of the zinc finger have the sequence DXSNXXR and the nucleotide sequence of the target subsite is GAC.

3. A zinc finger according to claim 1 wherein amino acids –1 through +6 of the zinc finger have the sequence RX(D/S)NXXR and the nucleotide sequence of the target subsite is GAG.

10 4. A zinc finger according to claim 1 wherein amino acids –1 through +6 of the zinc finger have the sequence TXGNXXR and the nucleotide sequence of the target subsite is GAT.

15 5. A zinc finger according to claim 1 wherein amino acids –1 through +6 of the zinc finger have the sequence (Q/T)XSNXXR and the nucleotide sequence of the target subsite is GAT.

6. A zinc finger according to claim 1 wherein amino acids –1 through +6 of the zinc finger have the sequence QXG(S/D)XXR and the nucleotide sequence of the target subsite is GCA.

20 7. A zinc finger according to claim 1 wherein amino acids –1 through +6 of the zinc finger have the sequence RXDEXXR and the nucleotide sequence of the target subsite is GCG.

25 8. A zinc finger according to claim 1 wherein amino acids –1 through +6 of the zinc finger have the sequence QXDSXXR and the nucleotide sequence of the target subsite is GCT.

9. A zinc finger according to claim 1 wherein amino acids -1 through +6 of the zinc finger have the sequence QX(G/A)HXXR and the nucleotide sequence of the target subsite is GGA.

10. A zinc finger according to claim 1 wherein amino acids -1 through +6 of the zinc finger have the sequence DXSHXXR and the nucleotide sequence of the target subsite is GGC.

11. A zinc finger according to claim 1 wherein amino acids -1 through +6 of the zinc finger have the sequence RXDHXXR and the nucleotide sequence of the target subsite is GGG.

12. A zinc finger according to claim 1 wherein amino acids -1 through +6 of the zinc finger have the sequence RXDAXXR and the nucleotide sequence of the target subsite is GTG.

13. A nucleic acid encoding a polypeptide wherein the polypeptide comprises a zinc finger according to claim 1.

14. A segment of a zinc finger comprising a sequence of seven contiguous amino acids as shown in any of Tables 1-5.

15. A nucleic acid encoding a polypeptide wherein the polypeptide comprises a segment of a zinc finger according to claim 14.

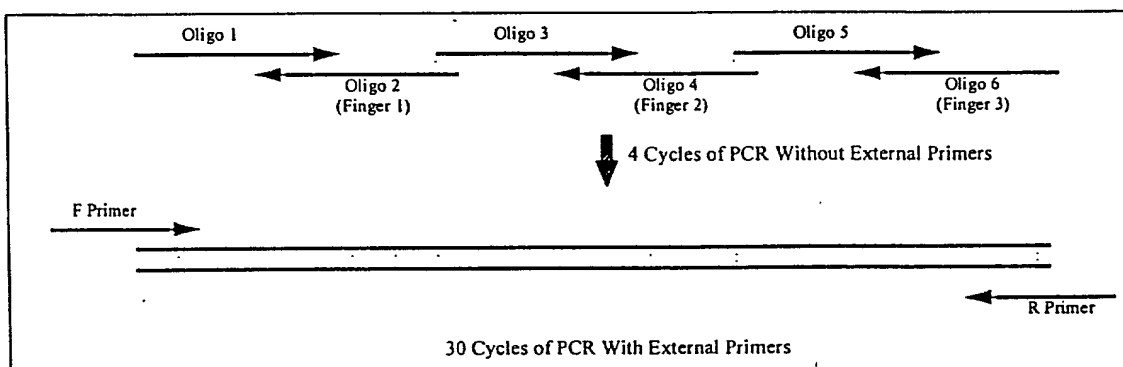
16. A zinc finger protein comprising first, second and third zinc fingers, wherein the zinc fingers comprise respectively first, second and third segments of seven contiguous amino acids as shown in a row of Tables 1-5.

17. A nucleic acid encoding a zinc finger protein according to claim 16.

CLOSURE

occurring zinc finger proteins, and consensus amino acid sequences for binding specificity are also provided.

Variable	Mean	Standard deviation	Minimum	Maximum
Age	34.5	10.2	21	55
Gender	0.5	0.5	0	1
Marital status	0.6	0.5	0	1
Education	12.5	1.5	9	16
Income	15.2	5.8	5	35
Health status	0.8	0.4	0	1
Employment status	0.7	0.5	0	1
Home ownership	0.9	0.3	0	1
Vehicle ownership	0.6	0.5	0	1
Life satisfaction	4.2	1.8	1	7
Subjective health	3.5	1.2	1	5
Life expectancy	78.5	3.2	70	85
Quality of life	5.8	2.1	1	9
Healthcare utilization	2.1	1.5	0	5
Health insurance	0.9	0.3	0	1
Healthcare costs	12.3	4.5	5	25
Healthcare access	0.7	0.4	0	1
Healthcare quality	4.5	1.5	1	7
Healthcare satisfaction	3.8	1.4	1	6
Healthcare equity	0.6	0.5	0	1
Healthcare efficiency	0.5	0.5	0	1
Healthcare effectiveness	0.4	0.4	0	1
Healthcare safety	0.3	0.3	0	1
Healthcare transparency	0.2	0.2	0	1
Healthcare accountability	0.1	0.1	0	1
Healthcare integrity	0.0	0.0	0	1
Healthcare trust	0.9	0.3	0	1
Healthcare confidence	0.8	0.4	0	1
Healthcare respect	0.7	0.5	0	1
Healthcare dignity	0.6	0.4	0	1
Healthcare autonomy	0.5	0.5	0	1
Healthcare privacy	0.4	0.4	0	1
Healthcare security	0.3	0.3	0	1
Healthcare justice	0.2	0.2	0	1
Healthcare fairness	0.1	0.1	0	1
Healthcare honesty	0.0	0.0	0	1
Healthcare openness	0.9	0.3	0	1
Healthcare communication	0.8	0.4	0	1
Healthcare collaboration	0.7	0.5	0	1
Healthcare partnership	0.6	0.4	0	1
Healthcare engagement	0.5	0.5	0	1
Healthcare involvement	0.4	0.4	0	1
Healthcare participation	0.3	0.3	0	1
Healthcare contribution	0.2	0.2	0	1
Healthcare commitment	0.1	0.1	0	1
Healthcare loyalty	0.0	0.0	0	1
Healthcare devotion	0.9	0.3	0	1
Healthcare dedication	0.8	0.4	0	1
Healthcare diligence	0.7	0.5	0	1
Healthcare industry	0.6	0.4	0	1
Healthcare profession	0.5	0.5	0	1
Healthcare career	0.4	0.4	0	1
Healthcare job	0.3	0.3	0	1
Healthcare work	0.2	0.2	0	1
Healthcare task	0.1	0.1	0	1
Healthcare role	0.0	0.0	0	1
Healthcare position	0.9	0.3	0	1
Healthcare status	0.8	0.4	0	1
Healthcare rank	0.7	0.5	0	1
Healthcare level	0.6	0.4	0	1
Healthcare degree	0.5	0.5	0	1
Healthcare grade	0.4	0.4	0	1
Healthcare class	0.3	0.3	0	1
Healthcare category	0.2	0.2	0	1
Healthcare group	0.1	0.1	0	1
Healthcare team	0.0	0.0	0	1
Healthcare unit	0.9	0.3	0	1
Healthcare department	0.8	0.4	0	1
Healthcare division	0.7	0.5	0	1
Healthcare branch	0.6	0.4	0	1
Healthcare section	0.5	0.5	0	1
Healthcare office	0.4	0.4	0	1
Healthcare center	0.3	0.3	0	1
Healthcare facility	0.2	0.2	0	1
Healthcare institution	0.1	0.1	0	1
Healthcare organization	0.0	0.0	0	1
Healthcare system	0.9	0.3	0	1
Healthcare network	0.8	0.4	0	1
Healthcare infrastructure	0.7	0.5	0	1
Healthcare framework	0.6	0.4	0	1
Healthcare structure	0.5	0.5	0	1
Healthcare model	0.4	0.4	0	1
Healthcare approach	0.3	0.3	0	1
Healthcare method	0.2	0.2	0	1
Health				



* PCR amplification scheme for production of ZFP-encoding synthetic genes.

FIG. 1

DECLARATION AND POWER OF ATTORNEY

As a below named inventor, I declare that:

My residence, post office address and citizenship are as stated below next to my name; I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural inventors are named below) of the subject matter which is claimed and for which a patent is sought on the invention entitled: **ZINC FINGER PROTEIN COMPOSITIONS** the specification of which X is attached hereto or was filed on as Application No. and was amended on (if applicable).

I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to patentability as defined in Title 37, Code of Federal Regulations, Section 1.56. I claim foreign priority benefits under Title 35, United States Code, Section 119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed.

Prior Foreign Application(s)

Country	Application No.	Date of Filing	Priority Claimed Under 35 USC 119

I hereby claim the benefit under Title 35, United States Code § 119(e) of any United States provisional application(s) listed below:

Application No.	Filing Date
60/126,238	March 24, 1999
60/126,239	March 24, 1999
60/146,596	July 30, 1999
60/146,615	July 30, 1999

I claim the benefit under Title 35, United States Code, Section 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, Section 112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, Section 1.56 which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

Application No.	Date of Filing	Status

POWER OF ATTORNEY: As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith.

Joe Liebeschuetz, Reg. No. 37,505
William M. Smith, Reg. No. 30,223

Send Correspondence to:
Joe Liebeschuetz
TOWNSEND and TOWNSEND and CREW LLP
Two Embarcadero Center, 8th Floor
San Francisco, California 94111-3834

Direct Telephone Calls to:
(Name, Reg. No., Telephone No.)
Name: Joe Liebeschuetz
Reg. No.: 37,505
Telephone: 650-326-2400

Full Name of Inventor 1:	Last Name: EISENBERG	First Name: STEPHEN	Middle Name or Initial: P.	
Residence & Citizenship:	City: Boulder	State/Foreign Country: CO	Country of Citizenship: United States	
Post Office Address:	Post Office Address: 5664 Pennsylvania	City: Boulder	State/Country: CO	Postal Code: 80303
Full Name of Inventor 2:	Last Name: CASE	First Name: CASEY	Middle Name or Initial: C.	
Residence & Citizenship:	City: San Mateo	State/Foreign Country: CA	Country of Citizenship: United States	
Post Office Address:	Post Office Address: 605 Barroilhet Avenue	City: San Mateo	State/Country: CA	Postal Code: 94402
Full Name of Inventor 3:	Last Name: COX III	First Name: GEORGE	Middle Name or Initial: N.	
Residence & Citizenship:	City: Louisville	State/Foreign Country: CO	Country of Citizenship: United States	
Post Office Address:	Post Office Address: 678 West Willow Street	City: Louisville	State/Country: CO	Postal Code: 80027
Full Name of Inventor 4:	Last Name: JAMIESON	First Name: ANDREW	Middle Name or Initial:	
Residence & Citizenship:	City: San Francisco	State/Foreign Country: CA	Country of Citizenship: United Kingdom	
Post Office Address:	Post Office Address: 2528 Sutter Street	City: San Francisco	State/Country: CA	Postal Code: 94115
Full Name of Inventor 5:	Last Name: REBAR	First Name: EDWARD	Middle Name or Initial: J.	
Residence & Citizenship:	City: Berkeley	State/Foreign Country: CA	Country of Citizenship: United States	
Post Office Address:	Post Office Address: 815 Regal Road	City: Berkeley	State/Country: CA	Postal Code: 94708
Full Name of Inventor 6:	Last Name: LUI	First Name: QIANG	Middle Name or Initial:	
Residence & Citizenship:	City: Foster City	State/Foreign Country: CA	Country of Citizenship: United States	
Post Office Address:	Post Office Address: 55 Williams Lane	City: Foster City	State/Country: CA	Postal Code: 94404

I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Signature of Inventor 1	Signature of Inventor 2	Signature of Inventor 3
STEPHEN P. EISENBERG	CASEY C. CASE	GEORGE N. COX III
Date	Date	Date
Signature of Inventor 4	Signature of Inventor 5	Signature of Inventor 6
ANDREW JAMIESON	EDWARD J. REBAR	QIANG LUI
Date	Date	Date

PA 3058856 v1